

[illegible]

XX The present sequence encodes a human inorganic pyrophosphatase designated HPYV. Nucleic acids encoding HPYV were first identified in inverte clone 768330 from the lung tissue cDNA library LUNGONC04 CC and tissue regeneration. CC

Sequence 1275 BP; 394 A; 235 C; 294 G; 351 T; 1 other;

Query Match	99.9%	Score 1274	DB 20	Length 1275
Best Local Similarity	100.0%	Pred. No: 0		
Matches 1275: Conservative	0	Mismatches	0	Indels 0
				Gaps 0

OY		1	CAAGATGTTGGGGCTCTTCMTCMTTGAACGCGGCCGCCTGGGGGTGGTGCTGCTCMTG	60
Db		1	cgaaggtcttuggggctctctccctccttgtaacgcgcgccgcgcgtgcgttcgtctctctg	60
OY		61	GCAAGCGGCGGCGCAGGACTCTCGGCACATGACGGCTTAGCACGAGAGCGCGCCG	120
Db		61	gcacgcgcgcgcgcgaaggactccgcacctatgtagcggtcttcagaccagaagcgcgcgc	120
OY		121	GCCCCCTGCTCCGTGAGTACCGAGGCTTCCTCCAAATAATGACAAGGACAAATTATCTCC	180
Db		121	gccctctcccctgaglaaccgaagtctctccccaataaaagagaagagacaatalatctcc	180
OY		181	ATTTCATGATTAATTCCAATTTATCATGATTAAGATGTGTTCACATGGTAGTTGAAGACC	240
Db		181	atttcatgatcttccaatttatitgtagtaagatgltgtttcacatigttagtgaagttacc	240
OY		241	ACCGTGTCTAATGCAAAAATGAGATGTCTACAAAGACCCCTTTAAACCTTATTAACA	300
Db		241	acgctgtctctaattgcaaaaaatgtagatctgtaccaaaagaccctttaaaccttaaca	300
OY		301	AGATGTAAAAAAGSAAACCTTCGCTATGTGTGGGAATTTCTCCCGTAAAGAATATAT	360
Db		301	agatgtgaaaaaaggaaaacttcgtctatgtttgcgaatttgtcccgtaataaagatatat	360
OY		361	CTGGAACATATGTTGCCATTCCTCAGACTTGGAAGAACCCAGGCGACAAATGATTAACATAC	420
Db		361	ctggaactatgtgtgcatactccctcagacttgggaagaccagcgacaaatgataaacatac	420
OY		421	TGGCTCTTGTGGTGACAAATACCAACATTAATGTGTGTGAATTGGAAATTGGAAGCAAGTATGTC	480
Db		421	tggctctgttgggtgacaaatbaccacaattgatlgtgtgtgaatttggaaagcaagttatgtc	480
OY		481	AAGAGTGAATTAATTTGGCGTGAAGTTCACAGCATTTGGCGATATTCAGCAAGGGGA	540
Db		481	aagagtggaataaatTTTTGGCTGGAAGTTCACAGCATTTGGCGATATTCAGCAAGGGGA	540
OY		541	AACGACGTGAAGATCTTTCCCATTAATATGTGAATGATCGATGACAGCCAAATTATATGA	600
Db		541	aacgcactggaagaatcatcttgccaattaatgtgtgattcctcgtgcagccaatataatga	600
OY		601	TATCAATGATGTCAACAGCGCTGAACCTGGCTACTTGAAGCTACTGTGACGTGTTTAC	660
Db		601	tatcaaagtatgtaaaaagcgtctgaacccttgctacttaagaacctgtgacgtgtttag	660
OY		661	AAGTATTAAGGTTCTCGATGTGAAGAAACACGAAATGAAGTTTGCCCTTAAATGACAATTAA	720
Db		661	aaggtataaagttcttcgtatgtgaaaaccgaanaatgagtttgcgtttaatgcagaatttaa	720
OY		721	AGATTAAGCATTTGCCATTGTATTAATTAATAAGCACATGACCATTTGGAAAGCATTTACT	780
Db		721	agataaaggacttgcgcattgatatlatlaaaagpacacatgaccatgtggaagaacattagt	780
OY		781	GACTTAACAAAACGATATGGAAGAAATCAATTCATGATTAATACAACTTTGCTAGAGCCC	840
Db		781	gacttaagaaaacgaatgtgaaaagaatcagtttgcatgaataacaactttgtctgaagccc	840
OY		841	CTTCAAGTGTGATCTGTATCTGCGACAGCCATTTGTGATGTCCTTTACACACCCCTGTGA	900
Db		841	cttcaagtgtgatacttcatatcttgcacaagccattgtgtagtcttaccacaccctgtga	900

QY	901	ATTCGCTGCACAGTACCAACACAGCGTGGATTAAGTGTTCATCACACAGAAAACTAATG	960
Db	901	atctgcctgcacagctaccaacagacgtgataagtggtctccatccacagaaaaactaatg	960
QY	961	AGATTCTCTCGAATTACACGTGATTTGGCTTAATCGTGTTCATCGTATGATATAGAG	1020
Db	961	agattctctcgaaatfacacgtgatattggcttaattcggttccatcgatcggatctatagaag	1020
QY	1021	TAAAGTAGTAGCTTTTCAAGGCTTTAAATTTGTAGACATCATTAACATTAAGTAAATTG	1080
Db	1021	taagaatagtagcttttccaagcttttaaatcttgtagaactcatctcaactaaagtaattc	1080
QY	1081	TGCTGTGACTAATCAATATATCTGAGANTGTATTCATCTAAAGCATTTTTCATATCTCA	1140
Db	1081	tgctgtgactaatccaatatactacagaagtgtatccatctaaagaattttcatatctcca	1140
QY	1141	ACTAAGTAACCTTTTACACACATGCTTAAATATATCAAGCGTTGCTATTTGGAAAGCATTT	1200
Db	1141	actaagataaacttttagcacaagtcttaataatcaaaagcagttgtaatttggaaagcactt	1200
QY	1201	GTGAAATAGATGTGCAGGAGGACACATATTGGATGTATATGTTACCATATGTTAGAAAT	1260
Db	1201	gtgataagatgtgcaagggagacatatgtgataatglttaccatatgtgtagaat	1260
QY	1261	AAAATTAATTTGCTG	1275
Db	1261	aaaatcatttgcgtg	1275

RESULT	3
AAS08596	
ID	AAS08596 standard; cDNA; 1277 BP.

AC AAS08596;

DT 24-OCT-2001 (first entry)

DE DNA encoding human pyrophosphatase (HPYP).

KM Pyrophosphatase; human; HPPV; antihypertensive; immunosuppressive; antineoplastic; antidiabetic; osteopathic; antirheumatic; antiarthritic; inflammatory disease; cancer; nerve; transplantation; prostate; prostate; anaemia; asthma; systemic lupus erythematosus; myasthenia gravis; diabetes mellitus; osteoporosis; glomerulonephritis; rheumatoid arthritis; osteoarthritis; scleroderma; ss.

Homo sapiens.

FH Key

ET

XX

XX

XX
00XX
31-00

PR 13-AU

РА (НАМК
РА (НТИТ

XX
NT
Work:XX
DP
WPT.:

DR P-PSD

PT New h

New human pyrophosphatase polypeptides and polynucleotides, useful for

Claim 5; Fig 1; 32pp; English.

XX Sequence 1277 BP; 394 A; 235 C; 294 G; 351 T; 3 other;
50

1 CAAGAGGTNGGGCTCTCTCCTTGTCAGTCGGCGCGCGGTGCGGGCTGGGAGGCTCTGTC

10. 5. 20

AAU00685
ID AAX00685 standard; DNA; 1351 BP

XX
DT 25-MAR-1999 (first entry)

XX human: secreted protein; gene

XX :

XX
PD 01-OCT-1998.

[illegible]

[illegible]

Db 721 aaggaacttgccatgatataataaagccatcagccattggaagacattagtgact 790
Oy 785 AAGAAACGAATGCAAGATGATGCTGATGATACACTTGTGTGAGAGCCCTTC 844
Db 791 aagaaacgaatggaagaaagatcatgttgatgttcaacttctgttgagccctcaaa 850
Oy 845 AAGTG 849
Db 851 gtgtg 855

RESULT 8
AAC10628
ID AAC10628 standard; cDNA: 660 BP.
XX
AC AAC10628;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 14703.
XX
KW Human: 5' EST: expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
FN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI: 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1: SEQ ID 14703; 71bp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 660 BP: 218 A; 112 C; 133 G; 197 T; 0 other;

Query Match 51.7%; Score 659; DB 21; Length 660;
Best Local Similarity 100.0%; Pred. NO. 4.5e-177;
Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 617 CGGCTGAACCTGGCTACTAGAACCTACTGTGACTGCTTTAGAGATTAAGCTTCT 676
Db 1 cggctgaacctggctactagaactactgtgactgtgtttagaagataagttctct 60
Oy 677 GATGGAACCAAGAAATGAGTTTGGCTTAATGCAAGATTAAAGATTAAGACTTTTGGC 736

Db 61 gatgaaacccaagaacatgagtttcgcttcaatgcaagattaaagacagacttgc 120
Oy 737 ATTGATATTATTAAGACATCATGATGATGATGATGATGATGATGATGATGAT 796
Db 121 attgattattttaaagaaagactcattgacattggaagattagtgactaaagaaagaa 180
Oy 797 GGAAGAAGATCATGTTGCTATGATGATGATGATGATGATGATGATGATGATGAT 856
Db 181 ggaagaaagaaatcagttgatgataatcaacttgcctgagagccctcaagtgtatcct 240
Oy 857 GATGCTGCCAGAGCCATGTTGATGATGATGATGATGATGATGATGATGATGAT 916
Db 241 gatgtgcagagagccattgtgagttgatttaccacaccccttgatctgcctgcacagta 300
Oy 917 CCAACAGACGTGATTAAGTGTGCTCATGATGATGATGATGATGATGATGATGAT 976
Db 301 ccaacagacgtgataagtgatgattcattcaccagaaacaaatgagattctcttggaata 360
Oy 977 CAAGCTGATATTGCTATCATGATGATGATGATGATGATGATGATGATGATGAT 1036
Db 361 caagctgataatgctcactcgtgtgtcattcattgagtgatgataagtaagtagtgcctt 420
Oy 1037 TCAAGCTTTAAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1096
Db 421 tcaagctttaaatattgtagagactcattcactaaagaaatctgcgtgactaattca 480
Oy 1097 ATATCTCAGATGTTATCATGATGATGATGATGATGATGATGATGATGATGAT 1156
Db 481 atatactagaatgttattcattcactaaagaaatctcactaaagaaatctcactaa 540
Oy 1157 GCACATGCTTAATTAATTCAGACAGTTCATGATGATGATGATGATGATGATGAT 1216
Db 541 gcaatgtcttaataataatacaagcagttgcatcttggaagtaactgtgataatgagta 600
Oy 1217 GGGAGACATATTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1275
Db 601 gggagacataatgtgagtaattgtacataatgttagaataaattatttgcgtg 659

RESULT 9
AAA78135/c
ID AAA78135 standard; cDNA: 578 BP.
XX
AC AAA78135;
XX
DT 14-NOV-2000 (first entry)
XX
DE cDNA encoding human colon tumour polypeptide, SEQ ID NO:422.
XX
KW Human colon tumour polypeptide; tumour antigen; cancer; vaccine;
KW immunotherapy; diagnosis; progression; ss.
XX
OS Homo sapiens.
XX
PN WO200037643-A2.
XX
PD 29-JUN-2000.
XX
PF 23-DEC-1999; 99WO-US30909.
XX
PR 23-DEC-1998; 98US-0221298.
XX
PR 02-JUL-1999; 99US-0347496.
XX
PR 22-SEP-1999; 99US-0401064.
XX
PR 19-NOV-1999; 99US-0444242.
XX
PR 02-DEC-1999; 99US-0454150.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Lodés MJ, Secretist H, Benson DR, Meagher MJ, Stolk J;
XX
DR Wang T, Yuguu J;
XX
WPI: 2000-442671/38.

XX New colon tumor polypeptides used to inhibit the development of cancer,
PT especially colon cancer, and for diagnosing and monitoring the
PT progression of the cancer -

Claim 1: Page 212; 229pp; English.

XX Sequences AAA77722-A78199 represent 478 cDNAs encoding proteins or
XX portions of proteins which are associated with human colon tumors.
CC The invention also specifically discloses 8 human colon tumor proteins
CC (AA11887-81104). The nucleic acids, the polypeptides they encode, and
CC antigen presenting cells (APCs), preferably dendritic cells) expressing
CC such polypeptides may be used in vaccines that target tumor cells,
CC especially colon tumor cells, thereby inhibiting the development of
CC cancer. T-cells specific for the polypeptide expressed by the APC are
CC used to remove tumor cells from biological samples, especially blood or
CC fractions thereof. The sample or the isolated T-cells specific for the
CC polypeptide can then be used to inhibit cancer development. CD4+ and/or
CC CD8+ T-cells from a patient may be incubated with a polypeptide or
CC nucleic acid of the invention, or an APC expressing such a polypeptide,
CC to cause the proliferation of specific T-cells. The T-cells can be
CC cloned and then administered back to the patient to inhibit cancer
CC development. Nucleic acids encoding the polypeptides and antibodies
CC against the polypeptides may be used to determine the expression level
CC of a tumor protein of the invention, and therefore to determine whether
CC cancer cells are present. Such diagnostic methods may also be used to
CC monitor the progression of a cancer by repeating the processes at time
CC intervals, and comparing the current result to previous results. The
CC present sequence represents a cDNA encoding a human colon tumor
XX polypeptide.

XX Sequence 578 BP; 155 A; 139 C; 99 G; 185 T; 0 other;

Query Match 44.0%; Score 560.6; DB 21; Length 578;
Best Local Similarity 99.1%; Pred. No. 3.6e-149; Indels 1; Gaps 1;
Matches 574; Conservative 0; Mismatches 4;

XX 337 TTTGTTCCGTTAAAGTATATCTGAACTATGTCATCCCTCAGACTTGGGAGA 396
DB 578 TTTGTTCCGTTAAAGTATATCTGAACTATGTCATCCCTCAGACTTGGGAGA 519
XX 397 CCCAGGCAATGATTAACATCTGCTGCTGTGTGACATGACCCCAATGATGTGTG 456
DB 518 CCCAGGCAATGATTAACATCTGCTGCTGTGTGACATGACCCCAATGATGTGTG 459
XX 457 TGAATTTGAGAGAGATGTCAGAGTGAATTAATTTGGCTGGAAGTTCTAGCAT 516
DB 458 TGAATTTGAGAGAGATGTCAGAGTGAATTAATTTGGCTGGAAGTTCTAGCAT 400
XX 517 ATTGGCTATGATGAGAGAGGGAACCGATGGAAGCATTCGATTAATGATGATGA 576
DB 399 ATTGGCTATGATGAGAGAGGGAACCGATGGAAGCATTCGATTAATGATGATGA 340
XX 577 TCCGATGACGCAATTAATATATGATTCATATGATGCAAAAGGCTGAGCTACTT 636
DB 339 TCCGATGACGCAATTAATATGATTCATATGATGCAAAAGGCTGAGCTACTT 280
XX 637 AGAAGTACTGTGACTGTTTGAAGAGTATAGTTCCTGATGGAAGCAAGAAATGA 696
DB 279 AGAAGTACTGTGACTGTTTGAAGAGTATAGTTCCTGATGGAAGCAAGAAATGA 220
XX 697 GTTTCGCTTATGACAGATTTAAAGATTAAGAGCTTGGCATGATTTTAAAGCAC 756
DB 219 GTTTCGCTTATGACAGATTTAAAGATTAAGAGCTTGGCATGATTTTAAAGCAC 160
XX 757 TCATGACCATTTGGAAGCATTTAGTACTAAGAAAACCAATGGAAGAGATCGTTGCAT 816
DB 159 TCATGACCATTTGGAAGCATTTAGTACTAAGAAAACCAATGGAAGAGATCGTTGCAT 100
XX 817 GAATGCACTTTGCTGAGAGCCCTTCAAGTGTGATCTGATGCTCCAGAGCATTTGT 876
DB 99 GAATGCACTTTGCTGAGAGCCCTTCAAGTGTGATCTGATGCTCCAGAGCATTTGT 40

XX 877 GGATGCTTTACCCACCACTGTCATCTGCTGCACACT 915
DB 39 GGATGCTTTACCCACCACTGTCATCTGCTGCACACT 1

RESULT 10
AA128873/C-
ID AA128873 standard; cDNA; 578 BP.
XX
XX AA128873;
XX
XX 12-OCT-2001 (first entry)
XX
XX Colon tumor related determined cDNA sequence for clone 31985.
XX
XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;
XX immunogenic; gene therapy; vaccine; colonic cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200149716-A2.
XX
XX 12-JUL-2001.
XX
XX 29-DEC-2000; 2000MO-US35596.
XX
XX 30-DEC-1999; 99US-0476296.
XX
XX 10-JAN-2000; 2000US-0480321.
XX
XX 15-FEB-2000; 2000US-0504629.
XX
XX 06-MAR-2000; 2000US-0519444.
XX
XX 19-MAY-2000; 2000US-0575251.
XX
XX 29-JUN-2000; 2000US-0609448.
XX
XX 28-AUG-2000; 2000US-0649811.

XX (CORI-) CORIXA CORP.
XX
XX Xu J, Lodes MJ, Secret H, Benson DR, Meagher MJ, Stolk JA;
XX Kling GE, Wang T, Jiang Y;
XX
XX WPI: 2001-441847/47.
XX
XX Colon tumor associated proteins and nucleic acids useful for the
XX prevention, diagnosis and treatment of colonic cancer -

PS Claim 2: Page 249; 472pp; English.

XX The present invention describes colon tumor associated proteins (I) and
XX the polynucleotides (II) that encode them. (I) have cytostatic activity.
XX (I) and (II) can be used in gene therapy, diagnosis and treatment of diseases
XX (II) may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate colon tumor associated protein (TCAP)
XX expression, such as colonic cancer. For example, (I) and (II) may be
XX used to treat disorders associated with decreased expression by
XX rectifying mutations or deletions in a patient's genome that affect the
XX activity of TCAPs by expressing inactive proteins or to supplement the
XX patient's own production of them. Additionally, (II) may be used to
XX produce the TCAP proteins, by inserting the nucleic acids into a host
XX cell culturing the cell to express the protein. (II) and its
XX complementary sequences may also be used as DNA probes in diagnostic
XX PCR and hybridisation assays to detect and
XX quantitate the presence of similar nucleic acids in samples, and
XX also be used as antigens in the production of antibodies against TCAPs
XX and in assays to identify modulators of TCAP expression and activity.
XX Anti-(I) antibodies and antagonists may also be used to down regulate
XX TCAP expression and activity. The anti-(I) antibodies may also be used
XX as diagnostic agents for detecting the presence of TCAPs in samples
XX (e.g. by enzyme linked immunosorbent assay (ELISA)). AA128460 to AA129512
XX given in the exemplification of the present invention.

XX Sequence 578 BP; 155 A; 139 C; 99 G; 185 T; 0 other;

Query Match 44.08; Score 560.6; DB 22; Length 578;
 Best Local Similarity 99.18; Pred. No. 3.6e-149;
 Matches 574; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 337 TTTGTCCTCCGATTAAGAGATATATCTGGAATATGTCGCCATTCCTGAGATTTGGGAAGA 396
 |||||||
 DB 578 TTTGTCCTCCGATTAAGAGATATATCTGGAATATGTCGCCATTCCTGAGATTTGGGAAGA 519
 |||||||
 QY 397 CCCAGGGCACAATGATAAACAATCTGCTGTTGTGTGACATATGACCAATTTGATGTGTG 456
 |||||||
 DB 518 CCCAGGGCACAATGATAAACAATCTGCTGTTGTGTGACATATGACCAATTTGATGTGTG 459
 |||||||
 QY 457 TGAATTTGGAAGCAAGATGTGCAAGAGTGAATAATTCGCCGTAAGTTTCAAGCAT 516
 |||||||
 DB 458 TGAATTTGGAAGCAAGATGTGCAAGAGTGAATAATTTCAATTTGAAGTTTCAAGCAT 400
 |||||||
 QY 517 ATTGGCTATGATGATGACGAAGGGAAGCAAGCACTGGAAGTCAATTCCTATTAATGTGATGA 576
 |||||||
 DB 399 ATTGGCTATGATGATGACGAAGGGAAGCAAGCACTGGAAGTCAATTCCTATTAATGTGATGA 340
 |||||||
 QY 577 TCTGTATGACACCAATTAATTAATGATATCAATGATGTCAAGGCTGAAGCCTGCTACTT 636
 |||||||
 DB 339 TCTGTATGACACCAATTAATTAATGATATCAATGATGTCAAGGCTGAAGCCTGCTACTT 280
 |||||||
 QY 637 AGAAGCTACTGTGACCTGCTTAAAGGTATAGGTTCCCTGATGGAAGCAAGCAAGCA 696
 |||||||
 DB 279 AGAAGCTACTGTGACCTGCTTAAAGGTATAGGTTCCCTGATGGAAGCAAGCAAGCA 220
 |||||||
 QY 697 GTTTCGCTTTAATGACAGATTTAAAGATTAAGAGACTTTGCCATTTATTTAATAAGACAC 756
 |||||||
 DB 219 GTTTCGCTTTAATGACAGATTTAAAGATTAAGAGACTTTGCCATTTATTTAATAAGACAC 160
 |||||||
 QY 757 TCATGACCATTTGGAAGCATTAATGATGACTAGAGAAAAGCAATGGAAGCAATTCAT 816
 |||||||
 DB 159 TCATGACCATTTGGAAGCATTAATGATGACTAGAGAAAAGCAATGGAAGCAATTCAT 100
 |||||||
 QY 817 GAATACAACTTTGCTGTGAGACCCCTCAAGTGTGATGCTGCTGCAAGCAATTCAT 876
 |||||||
 DB 99 GAATACAACTTTGCTGTGAGACCCCTCAAGTGTGATGCTGCTGCAAGCAATTCAT 40
 |||||||
 QY 877 GGATGCTTTTACCACACCCTGTGATCTGCTGCACACT 915
 |||||||
 DB 39 GGATGCTTTTACCACACCCTGTGATCTGCTGCACACT 1

RESULT 11
 AAI13542/c
 ID AAI13542 standard; DNA; 411 BP.
 AC AAI13542;
 XX
 XX 12-OCT-2001 (first entry)
 DE Probe #3475 for gene expression analysis in human cervical cell sample.
 XX
 KW Probe: human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; ss.
 OS Homo sapiens.
 XX
 XX WO200157278-A2.
 PN
 XX
 PD 09-AUG-2001.
 XX
 XX 30-JAN-2001; 2001WO-US00670.
 PF
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488901/53.
 XX
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 analyzing gene expression in human cervical epithelial cells -
 XX
 PS Claim 25; SEQ ID NO 3475; 487bp; English.

CC The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human Hela cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

Sequence: 411 BP; 106 A; 101 C; 67 G; 137 T; 0 other;

Query Match 30.48; Score 387; DB 22; Length 411;
 Best Local Similarity 96.48; Pred. No. 5.8e-100;
 Matches 396; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 273 CAAAGACCCCTTTAAACCTTTAAACAGATGGAAGAAAAGCAAACTGCTATGTTG 332
 |||||||
 DB 411 CAAAGACCCCTTTAAACCCCTTTAAACAGATGGAAGAAAAGCAAACTGCTATGTTG 352
 |||||||
 QY 333 CGAATTTGTTCCCTATTAAGATATATCTGGAATATGTCGCCATTCCTGAGACTTGGG 392
 |||||||
 DB 351 CGAATTTGTTCCCTATTAAGATATATCTGGAATATGTCGCCATTCCTGAGACTTGGG 292
 |||||||
 QY 393 AAGACCCAGGCGACATGATTAACATCTGCTGTTGTGTGACATGACCAATTCATG 452
 |||||||
 DB 291 AAGACCCAGGCGACATGATTAACATCTGCTGTTGTGTGACATGACCAATTCATG 232
 |||||||
 QY 453 TGTGTGAATTTGAAGCAAGTATGTCGAAGGGAATTAATGCGCTGAAGTTGAG 512
 |||||||
 DB 231 TGTGTGAATTTGAAGCAAGTATGTCGAAGGGAATTAATGCGCTGAAGTTGAG 172
 |||||||
 QY 513 GCATATTTGCTATGATTTGACGAAGGGGAACCGACTGGAAGTCAATTCCTAATGTGG 572
 |||||||
 DB 171 GCATATTTGCTATGATTTGACGAAGGGGAACCGACTGGAAGTCAATTCCTAATGTGG 112
 |||||||
 QY 573 ATGATCTGATGACACCAATTAATGATATCATGATGATCAAGCGCTGAAGCTGGCT 632
 |||||||
 DB 111 ATGATCTGATGACACCAATTAATGATATCAATGATGATCAAGCGCTGAAGCTGGCT 52
 |||||||
 QY 633 ACTTAGAGCTACTGCTGACTGCTTTAGAGGTATTAAGTCTCGATGGA 683
 |||||||
 DB 51 ACTTAGAGCTACTGCTGACTGCTTTAGAGGTATTAAGTCTCGATGGA 1

RESULT 12
 AAI34904/c
 ID AAI34904 standard; DNA; 411 BP.
 AC AAI34904;
 XX
 XX 17-OCT-2001 (first entry)
 DE Probe #3350 used to measure gene expression in human placenta sample.
 XX
 KW Probe: microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.

XX OS Homo sapiens.
XX PN W0200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-48897/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta
XX PS Claim 25; SEQ ID NO 3590; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENPs).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders.
XX SQ Sequence 411 BP; 106 A; 101 C; 67 G; 137 T; 0 other;
Query Match 30.4%; Score 387; DB 22; Length 411;
Best Local Similarity 96.4%; Pred. No. 5.8e-100; Indels 0; Gaps 0;
Matches 396; Conservative 0; Mismatches 15;
XX 273 CAAGGACCCCTTAAACCCCTATTAAACAGATGTGAAAAAGGAAAACTTCGCTATGTTG 332
XX 411 CAAGGACCCCTTAAACCCCTATTAAACAGATGTGAAAAAGGAAAACTTCGCTATGTTG 352
XX 333 CGAATTTGTTCCCGTATTAAGATATATCTGGAATGTCCTCCCTGAGCTGGG 392
XX 351 CGAATTTGTTCCCGTATTAAGATATATCTGGAATGTCCTCCCTGAGCTGGG 292
XX 393 AAGACCCAGGCGACAAATGATTAACACTACTGCTGTGTGAGACATGACCAATTGATG 452
XX 291 AAGACCCAGGCGACAAATGATTAACACTACTGCTGTGTGAGACATGACCAATTGATG 232
XX 453 TGTGTGAATTTGGAAGCAAGGTATGTGCAAGAGTGAATAATTTGGCTGAAGTTCTAG 512
XX 231 TGTGTGAATTTGGAAGCAAGGTATGTGCAAGAGTGAATAATTTGGCTGAAGTTCTAG 172
XX 513 GCATATTTGCTATGATTTGCAAGAGGGAACCGACTGGAAGTCAATTCCTATATGTTG 572
XX 171 GCATATTTGCTATGATTTGCAAGAGGGAACCGACTGGAAGTCAATTCCTATATGTTG 112
XX 573 ATGATTCCTGATGAGCCAAATTAATGATATGCAATGATGTCAAGAGGCTGAAGCTGGCT 632
XX 111 ATGATTCCTGATGAGCCAAATTAATGATATGCAATGATGTCAAGAGGCTGAAGCTGGCT 52
XX 633 ACTTAGACGACTGTGAGCTGTTAGAGAGTATTAAGTTCTCTGATGAA 683
XX 51 ACTTAGACGACTGTGAGCTGTTAGAGAGTATTAAGTTCTCTGATGAA 1
Db
RESULT 13
AA103431/C
ID AA103431 standard; DNA: 411 BP.

XX AC AA103431;
XX DT 09-OCT-2001 (first entry)
XX DE Probe #3422 used to measure gene expression in human breast sample.
XX KW Probe: human; breast disease; breast cancer; development disorder; ss;
XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX OS Homo sapiens.
XX PN W0200157270-A2.
XX PD 09-AUG-2001.
XX PF 29-JAN-2001; 2001WO-US00661.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-476286/51.
XX PT Novel single exon nucleic acid probe used to measuring gene expression
XX PT in a human breast
XX PS Claim 25; SEQ ID NO 3422; 322pp; English.
XX CC The present invention relates to novel single exon nucleic acid probes.
XX CC The present sequence is one such probe. The probes are useful for
XX CC measuring human gene expression in a human breast sample, where the probe
XX CC hybridises at high stringency to a nucleic acid expressed in the human
XX CC breast. The probes are useful for predicting, diagnosing, grading,
XX CC staging, monitoring and prognosing diseases of the human breast,
XX CC particularly those diseases with polygenic aetiology. The diseases
XX CC include: breast cancer, disorders of development, inflammatory diseases
XX CC of the breast, fibrocystic changes, proliferative breast disease and
XX CC non-carcinoma tumours.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 411 BP; 106 A; 101 C; 67 G; 137 T; 0 other;
Query Match 30.4%; Score 387; DB 22; Length 411;
Best Local Similarity 96.4%; Pred. No. 5.8e-100; Indels 0; Gaps 0;
Matches 396; Conservative 0; Mismatches 15;
XX 273 CAAGGACCCCTTAAACCCCTATTAAACAGATGTGAAAAAGGAAAACTTCGCTATGTTG 332
XX 411 CAAGGACCCCTTAAACCCCTATTAAACAGATGTGAAAAAGGAAAACTTCGCTATGTTG 352
XX 333 CGAATTTGTTCCCGTATTAAGATATATCTGGAATGTCCTCCCTGAGCTGGG 392
XX 351 CGAATTTGTTCCCGTATTAAGATATATCTGGAATGTCCTCCCTGAGCTGGG 292
XX 393 AAGACCCAGGCGACAAATGATTAACACTACTGCTGTGTGAGACATGACCAATTGATG 452
XX 291 AAGACCCAGGCGACAAATGATTAACACTACTGCTGTGTGAGACATGACCAATTGATG 232
XX 453 TGTGTGAATTTGGAAGCAAGGTATGTGCAAGAGTGAATAATTTGGCTGAAGTTCTAG 512
XX 231 TGTGTGAATTTGGAAGCAAGGTATGTGCAAGAGTGAATAATTTGGCTGAAGTTCTAG 172
Db

QY 513 GCATATTGGCTATGATGACGAAGGGAACCGACTGGAAGTCAATTCATTAATGCG 572
 Db 171 GCATATTGGCTATGATGACGAAGGGAACCGACTGGAAGTCAATTCATTAATGCG 112
 QY 573 ATGATTCCTGATGACGCAATTAATATGATATCAATGATGATCAAAAGGCTGAACCTGCT 632
 Db 111 ATGATTCCTGATGACGCAATTAATATGATATCAATGATGATCAAAAGGCTGAACCTGCT 52
 QY 633 ACTTAGAAGCTACTGCTGACGCTTTAGAGGTATAGCTCCCTGATGGA 683
 Db 51 ACTTAGAAGCTACTGCTGACGCTTTAGAGGTATAGCTCCCTGATGGA 1

RESULT 14

AAT21836
 ID AAT21836 standard; cDNA to mRNA; 416 BP.

AC AAT21836;

DT 01-AUG-1996 (first entry)

DE Human gene signature HUMS03377.

KW Gene signature: messenger RNA; mRNA; relative abundance; frequency;
 KW human: cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.

OS Homo sapiens.

PN W09514772-A1.

PD 01-JUN-1995.

PF 11-NOV-1994; 94WO-JP01916.

PR 12-NOV-1993; 93JP-0355504.

PA (MATS/) MATSUBARA K.
 PA (OKUB/) OKUBO K.

PI Matsubara K, Okubo K;

DR WPI; 1995-206931/27.

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.

PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues

PS Claim 1; Page 977; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in AAT19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.

SO Sequence 416 BP; 128 A; 77 C; 77 G; 128 T; 6 other;

Query Match 29.9%; Score 381.2; DB 16; Length 416;
 Best Local Similarity 97.6%; Pred. No. 2.6e-98;

Matches 405; Conservative 0; Mismatches 8; Indels 2; Gaps 2;
 QY 851 GATCCGATGCTGCGACAGCCATGTGGATGCTTACACACCGCTGGATTCGCGC 910
 Db 1 gatccgatgctgccaagccatgtggaagcttaccacacccgtggaatcgccgc 60
 QY 911 ACAGTACCAACAGAGCTGATTAAGTGTTCATCAGCAGAAATATAGATTTCTCT 970
 Db 61 acagtaccacaagcgtggaataagtggttcacatcacagaanaactaagatctct 120
 QY 971 GGAATACAGCTGATATTGCTACATCGTGTTCATCTGGATGATTAAGTAAGTAGT 1030
 Db 121 ggaatacagctgataatgtgctacacatcglttcacatcgtatggaagaagtagt 180
 QY 1031 AGCTTTCAAGCTTTAAATTTAGAACATCATCTAATTAAGTAATTCGCTGACT 1090
 Db 181 agctttcaagctttaaatltgtaagaactcatctaactaaatltcgctgtagct 240
 QY 1091 AATCCAATATACAGCAATGTTATTCATCTAAGCATTTTTCATATCTCAACTAAGATTA 1150
 Db 241 aatccaatatactcaagaatgtatccatctaaagcatlttcacatcactaagataa 300
 QY 1151 CTTTACGACATGCTTAATATCAAAAGCAGTTGTCATTGGAGCTGATGATAGAT 1210
 Db 301 cttttagcacatgctttaataatacaagcagtgctcattggaagctgtagaatagat 360
 QY 1211 GTGCAAGG-GGAGCACATATTGATGTATATG-TTACCATATGTTAGCAATTA 1263
 Db 361 gtgcaagggcggacacatcgtgntgataatgttncatgtatggaagataaa 415

RESULT 15

AAX89287
 ID AAX89287 standard; DNA; 1198 BP.

AC AAX89287;

DT 21-SEP-1999 (first entry)

DE Human regulatory protein HRGP-3 encoding DNA.

KW Regulatory protein; HRGP; human; cell proliferation; immune response;
 KW cancer; adenocarcinoma; leukemia; lymphoma; melanoma; myeloma; sarcoma;
 KW tetracarcinoma; Addison's disease; adult respiratory distress syndrome;
 KW allergy; anemia; asthma; atherosclerosis; bronchitis; cholelithiasis;
 KW Crohn's disease; ulcerative colitis; atopic dermatitis; dermatomyositis;
 KW diabetes mellitus; emphysema; atrophic gastritis; glomerulonephritis;
 KW gout; Grave's disease; hyperlipidemia; irritable bowel syndrome; AIDS;
 KW lupus erythematosus; multiple sclerosis; myasthenia gravis; infection;
 KW osteoarthritis; osteoporosis; pancreatitis; polymyositis; hemodialysis;
 KW arthritis; scleroderma; Sjogren's syndrome; and autoimmune thyroiditis;
 KW infection; trauma; ss.

OS Homo sapiens.

PN W09933870-A2.

PD 08-JUL-1999.

PF 22-DEC-1998; 98WO-US27471.

PR 31-DEC-1997; 97US-0001403.

PA (INCY-) INCYTE PHARM INC.

PI Au-Young J, Bandman O, Corley NC, Guegler KJ, Hillman JL;

DR Lal P, Shah P, Tang YT, Yue H;

DR WPI; 1999-430229/36.

PT New human regulatory proteins, useful for diagnosing, preventing and

PT treating disorders associated with expression of regulatory proteins

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2002, 18:52:37 ; Search time 1849.3 Seconds
(without alignments)

11373.981 Million cell updates/sec

Title: US-09-415-540-2

Perfect score: 1275
Sequence: 1 CAAGAGGTGNGGGCTCTCT.....GAATTAATTAATTTGCTG 1275Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

```
GenBml: *
1:  gb_da: *
2:  gb_htg: *
3:  gb_in: *
4:  gb_on: *
5:  gb_ov: *
6:  gb_pat: *
7:  gb_ph: *
8:  gb_pl: *
9:  gb_pr: *
10: gb_ro: *
11:  gb_sts: *
12:  gb_sy: *
13:  gb_un: *
14:  gb_vl: *
15:  em_ba: *
16:  em_fun: *
17:  em_hum: *
18:  em_in: *
19:  em_om: *
20:  em_or: *
21:  em_ov: *
22:  em_pat: *
23:  em_ph: *
24:  em_pl: *
25:  em_ro: *
26:  em_sts: *
27:  em_sy: *
28:  em_un: *
29:  em_vl: *
30:  em_htgo_hum: *
31:  em_htgo_inv: *
32:  em_htgo_rod: *
33:  em_htg_hum: *
34:  em_htg_inv: *
35:  em_htg_rod: *
36:  em_htg_other: *
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1274	99.9	1275	6	AR061907 Sequence
2	1274	99.9	1275	6	AR084921 Sequence
3	1266.8	99.4	1303	6	AX015063 Sequence
4	1265.8	99.3	1301	9	BC001022 Homo sapi
5	1264	99.1	1282	9	AF119665 Homo sapi
6	1228.8	96.4	1243	9	AF154065 Homo sapi
7	1217	95.5	1239	9	AB026723 Homo sapi
8	1185.4	93.0	1204	9	AF217186 Homo sapi
9	1179.4	92.5	1200	6	AX018061 Sequence
10	1123.8	88.1	167012	2	AC008429 Homo sapi
11	1123.8	88.1	213025	2	AC021159 Homo sapi
12	1079.2	84.6	155645	2	AC011012 Homo sapi
13	917.2	71.9	1266	4	BOV108211 Bovine inor
14	841.2	66.0	846	9	AF108211 Homo sapi
15	835.8	65.6	1243	10	BC010468 Mus muscu
16	799.2	62.7	856	6	AX014862 Sequence
17	560.6	44.0	578	6	AX192855 Sequence
18	546.4	42.9	70575	2	AC025103 Homo sapi
19	391.2	30.7	1088	5	AF174582 Torpedo m
20	380.2	30.6	70575	2	AC025103 Homo sapi
21	375.8	29.5	1202	6	AX011614 Sequence
22	375.8	29.5	1214	9	AF217187 Homo sapi
23	375.8	29.5	1706	9	AX000466 Homo sapi
24	368.8	28.9	1188	6	AX086197 Sequence
25	368.8	28.9	1188	9	HSMB01625 Homo sapi
26	350	27.5	165756	2	AL355138 Homo sapi
27	350	27.5	215049	2	AC067749 Homo sapi
28	339.2	26.6	1225	10	BC011417 Mus muscu
29	238.2	18.7	1115	9	AB026722 Homo sapi
30	236.6	18.6	938	11	CNS06P1E Drosophi
31	235.4	18.5	873	3	AF085601 Drosophi
32	229.6	18.0	1612	8	SCPPAG Scyllar
33	229.6	18.0	2411	8	SCYBR011C Scyllar
34	226.6	17.8	2887	8	KLTRPLOC KLTRP
35	223	17.5	2326	8	PETPR1877 Homo sapi
36	221.4	17.4	184180	9	HS1042K10 Human PA
37	215	16.9	2759	8	ZBA309279 Zyxosacch
38	212.6	16.7	720	8	CNS01DAS Caenorhabd
39	208	16.3	37855	3	CEC47E12 CeC47E12
40	202.2	15.9	357	6	AX014834 Sequence
41	201.2	15.8	636	8	CNS01A30 Cns01A30
42	191.8	15.0	1366	8	SPPAG Scyllar
43	191.8	15.0	38391	8	SPAC23C11 Arabidops
44	190.8	15.0	911	8	ATH252210 Arabidops
45	190.8	15.0	1161	8	AY045945 Arabidops

ALIGNMENTS

RESULT	1	PAT	29-SEP-1999
AR061907	AR061907	1275 bp	DNA
LOCUS	Sequence 2 from patent US 5843665.		
DEFINITION	AR061907		
ACCESSION	AR061907.1	GI:5989598	
VERSION			
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1275)		
AUTHORS	Hawkins, P.R. and Hillman, J.L.		
TITLE	Human pyrophosphatase		
JOURNAL	Patent: US 5843665-A 2 01-DEC-1998;		
FEATURES	Location/Qualifiers		
source	1..1275		
BASE COUNT	394 a 235 c 294 g 351 t 1 others		
ORIGIN	/organism="unknown"		

QY	1021	TAAAGTAGAGCTTTTCAAGAGCTTTAAATTTGTAGAACTACTTAACTAAGTAATTC	1080
Db	1021	TAAAGTAGAGCTTTTCAAGAGCTTTAAATTTGTAGAACTACTTAACTAAGTAATTC	1080
QY	1081	TGCGTGTACTAATCCAAATATATCTCAAAATGTTATATCATTTTTCATATCTCA	1140
Db	1081	TGCGTGTACTAATCCAAATATATCTCAAAATGTTATATCATTTTTCATATCTCA	1140
QY	1141	ACTAGATAACTTTTACGACATGCTTAAATATCAAGCAGTTGTCATTTTGGAAAGTCACTT	1200
Db	1141	ACTAGATAACTTTTACGACATGCTTAAATATCAAGCAGTTGTCATTTTGGAAAGTCACTT	1200
QY	1201	GTGAATAGATGTGCAAGGGGAGCACAATTGATGATGTATGTATCCATATGTTAGAAAT	1260
Db	1201	GTGAATAGATGTGCAAGGGGAGCACAATTGATGATGTATGTATCCATATGTTAGAAAT	1260
QY	1261	AAATATATTTTGGTCTG 1275	
Db	1261	AAATATATTTTGGTCTG 1275	
RESULT	2		
LOCUS	AR084921	1275 bp	DNA
DEFINITION	Sequence 2 from patent US 5981232.		PAT
ACCESSION	AR084921		
VERSION	AR084921.1	GI:10011692	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1. (bases 1 to 1275)		
AUTHORS	Hawkins, P.R. and Hillman, J.L.		
TITLE	Human pyrophosphatase		
JOURNAL	Patent: US 5981232-A 2, 09-NOV-1999;		
FEATURES	Location/Qualifiers		
source	1..1275		
BASE COUNT	394 a 235 c 294 g 351 t	1 others	
ORIGIN			
Query Match	99.9%; Score 1274; DB 6; Length 1275;		
Best Local Similarity	100.0%; Pred. No. 1.3e-284;		
Matches 1275; Conservative	0; Mismatches 0; Indels 0; Gaps		
QY	1	CAAGAGTTNGGGGCTG	60
Db	1	CAAGAGTTNGGGGCTG	60
QY	61	GCACGGCGGGGCGGCGAGACTCCGGCAGTATGAGCGGCTTCAGCACCGAGAGAGCGCGCGG	120
Db	61	GCACGGCGGGGCGGCGAGACTCCGGCAGTATGAGCGGCTTCAGCACCGAGAGAGCGCGCGG	120
QY	121	GCCTTCCTCCCTGGAGTACCGAGTCTTCTCTCAAAATGAGAAAGGACAAATATATATCTCC	180
Db	121	GCCTTCCTCCCTGGAGTACCGAGTCTTCTCTCAAAATGAGAAAGGACAAATATATATCTCC	180
QY	181	ATTTCATGATTTTCCATTTATGTCAGATAGAGATGTGTTTCATGTGATGTTGAAGTACC	240
Db	181	ATTTCATGATTTTCCATTTATGTCAGATAGAGATGTGTTTCATGTGATGTTGAAGTACC	240
QY	241	ACCGTGTCTAATCCAAAAATGAGATGTCTACAAAGAGACCCCTTTAAACCTATTAACA	300
Db	241	ACCGTGTCTAATCCAAAAATGAGATGTCTACAAAGAGACCCCTTTAAACCTATTAACA	300
QY	301	AGATGTGAAAAAGAAACCTGCTATGTGTGCAATTTTCTCCGTATTAAGATATAT	360
Db	301	AGATGTGAAAAAGAAACCTGCTATGTGTGCAATTTTCTCCGTATTAAGATATAT	360
QY	361	CTGGAACTATGTGTGCATCCCTCAGACTTGGGAAGACCCAGGGGACAAATATTAACATAC	420
Db	361	CTGGAACTATGTGTGCATCCCTCAGACTTGGGAAGACCCAGGGGACAAATATTAACATAC	420

[illegible]

RESULT	5
LOCUS	AF119665
DEFINITION	AF119665 Homo sapiens inorganic pyrophosphatase mRNA, complete cds.
ACCESSION	AF119665
VERSION	AF119665.1
KEYWORDS	GI:6563255
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 1282) Gu,Y., Peng,Y., Li,Y., Fu,S., Gu,G., Zhang,L., Zhang,C., Yu,Y., Han,Z., Wang,Y., Chen,Z., and Fu,G.
TITLE	A novel gene expressed in the human adrenal gland
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1282)
AUTHORS	Huang,C., Zhang,C., Wu,T., Peng,Y., Gu,Y., Zhang,L., Jiang,C., Li,Y., Han,Z., Wang,Y., Chen,Z., and Fu,G.
TITLE	Direct Submission
JOURNAL	Submitted (12-JAN-1999) Chinese National Human Genome Center at Shanghai, Guo Shoujing Rd. 351, Zhangjiang Hi-Tech Park, Pudong

FEATURES	Shanghai 201203, P. R. China
Source	Location/Qualifiers
CDS	1. 1282 /organism="Homo sapiens" /db_xref="taxon:9606" /lssue_type="adrenal gland" 78. 947 /note="IOPPP" /codon_start=1 /evidence=not-experimental /product="inorganic pyrophosphatase" /protein_id="AAPI7222.1" /db_xref="GI:5653256"
BASE COUNT	409 a 234 c 290 g 349 t
ORIGIN	PTVDKMFHHOKN
Query Match	99.1%; Score 1264; DB 9; Length 1282;
Best Local Similarity	100.0%; Pred. No. 2,7e-282;
Matches 1264; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Oy	12 GGGGCTCTCCCTTCGTCAGTCGGCGCGCGGTGGGCTGGTGGCTCTGTGGCAGCGGCGGC 71
Db	1 GGGCTCTCTCTTCGTCAGTCGGCGCGCGGTGGGCTGGTGGCTCTGTGGCAGCGGCGGC 60
Oy	72 GGCAGGACTCGGCGCAGTATGAGCGGCTTCAGCACCAGAGAGCGCGCGCCCTTCCTCC 131
Db	61 GGCAGGACTCGGCGCAGTATGAGCGGCTTCAGCACCAGAGAGCGCGCGCCCTTCCTCC 120
Oy	132 TGGAGTACCGAGTCTCTCTCAAAATGAGAAAGACATATATCTCTCATTTGATGATA 191
Db	121 TGGAGTACCGAGTCTCTCTCAAAATGAGAAAGACATATATCTCTCATTTGATGATA 180
Oy	192 TTCCAAATTATGACATTAAGGATGTGTTTTCACATGATGATTAAGTACACGCTGGTCTA 251
Db	181 TTCCAAATTATGACATTAAGGATGTGTTTTCACATGATGATTAAGTACACGCTGGTCTA 240
Oy	252 ATGCAAAAATGAGATTCCTACAAAAGACCCCTTTAAACCCATTAAACAGATGTGAANA 311
Db	241 ATGCAAAAATGAGATTCCTACAAAAGACCCCTTTAAACCCATTAAACAGATGTGAANA 300
Oy	312 AAGGAAACCTTGGCTATGTTGGCAATTTGTTCCGCTATTAAGGATATCTTGAACATATG 371
Db	301 AAGGAAACCTTGGCTATGTTGGCAATTTGTTCCGCTATTAAGGATATCTTGAACATATG 360
Oy	372 GTGGCATCCCTCAGACTTGGGAAGACCCAGGACAAATGATTAACATCTGGCTGTGTG 431
Db	361 GTGGCATCCCTCAGACTTGGGAAGACCCAGGACAAATGATTAACATCTGGCTGTGTG 420
Oy	432 GTGACAATGACCAATGATGTGTGAANAATGGAAGCAAGGTATGTCCAAAGGTGAAA 491
Db	421 GTGACAATGACCAATGATGTGTGAANAATGGAAGCAAGGTATGTCCAAAGGTGAAA 480
Oy	492 TAATTGGGCTGAAGTTCTAGGCATATTGGCTATGATTTGACGAGGGGAAACCGACTGGA 551
Db	481 TAATTGGGCTGAAGTTCTAGGCATATTGGCTATGATTTGACGAGGGGAAACCGACTGGA 540
Oy	552 AAGTATTGGCTTATGTGATGATTCGATGACGAGCCAAATATATATGATCAATGATG 611
Db	541 AAGTATTGGCTTATGTGATGATTCGATGACGAGCCAAATATATATGATCAATGATG 600
Oy	612 TCAAAAGCGGTGAACCTGGCTACTTAGAAGCTACTGTGACTGCTTTAGAAAGTATAAGG 671
Db	601 TCAAAAGCGGTGAACCTGGCTACTTAGAAGCTACTGTGACTGCTTTAGAAAGTATAAGG 660
Oy	672 TTTCGATGGAAGAACCAAGAAATGAGTTTGCTTTAATGCAAGATTTTAAAGTAAAGACT 731
Db	661 TTTCGATGGAAGAACCAAGAAATGAGTTTGCTTTAATGCAAGATTTTAAAGTAAAGACT 650

Query Match	93.0%	Score 1185.4	DB 9	Length 1204
Best Local Similarity	99.9%	Pred. No. 4.2e-264		
Matches 1186	Conservative	0	Mismatches 1	Indels 0
			Gaps 0	

0y 89 ATGACGGCTTACACACCAGAGACGGCGCCCTTCTCCCTGGAGTACGAGTCTTC 148

Db	1	ATGAGCGCTTCACGACACCAGAGGCGCGCGGCCCTTCCTCCCTGAGTACCGAGCTTTC	60
Qy	149	CTCAAAATGAGAAAGACAAATATATATCTCCATTTCAATGATATCCAAATTAATGACGAT	208
Db	61	CTCAAAATGAGAAAGACAAATATATATCTCCATTTCAATGATATCCAAATTAATGACGAT	120
Qy	209	AAGATGTGTTTCAATGAGTAGTGAAGTACACAGCGTGGTCAATAGCAAAAATGAGATT	268
Db	121	AAGATGTGTTTCAATGAGTAGTGAAGTACACAGCGTGGTCAATAGCAAAAATGAGATT	180
Qy	269	GCTACAAAGACCCCTTTAAACCCCTATTTAAACAAGATGTGAAAAAAGAAACCTTGCAT	328
Db	181	GCTACAAAGACCCCTTTAAACCCCTATTTAAACAAGATGTGAAAAAAGAAACCTTGCAT	240
Qy	329	GTTGCGAATTTGTCCCGTAAAGATATATGTGAATATAGTGCGCATCCCTCAGACT	388
Db	241	GTTGCGAATTTGTCCCGTAAAGATATATGTGAATATAGTGCGCATCCCTCAGACT	300
Qy	389	TGGGAAGACCCAGGCGACAAATGATAAACATACTGGCTGTTGTGTGCATATGCCAAAT	448
Db	301	TGGGAAGACCCAGGCGACAAATGATAAACATACTGGCTGTTGTGTGCATATGCCAAAT	360
Qy	449	GATGTGTGTGAATTTGGAGACAGGTATGTGCAAGAGGGTAAATATATGGGTGAAGTT	508
Db	361	GATGTGTGTGAATTTGGAGACAGGTATGTGCAAGAGGGTAAATATATGGGTGAAGTT	420
Qy	509	CTAGCATATTGGCGTATGATTTAGCAAGGGGAAACCGACTGGAAMGCTATTTGCCATTAT	568
Db	421	CTAGCATATTGGCGTATGATTTAGCAAGGGGAAACCGACTGGAAMGCTATTTGCCATTAT	480
Qy	569	GTTGATGTATCTGTATGACGCCAATTAATATGATATCAATATGTCAAAGGCTGAACCT	628
Db	481	GTTGATGTATCTGTATGACGCCAATTAATATGATATCAATATGTCAAAGGCTGAACCT	540
Qy	629	GGCTACTTAGAAGCTACTGCTGAGCTGGTTTGAAGGATATAGGTTCTCTGATGAAAAACA	688
Db	541	GGCTACTTAGAAGCTACTGCTGAGCTGGTTTGAAGGATATAGGTTCTCTGATGAAAAACA	600
Qy	689	GAAATATGATTTGGCTTTAAATGACGAATTTAAAGATAGAGACTTTGCCATGATATTAT	748
Db	601	GAAATATGATTTGGCTTTAAATGACGAATTTAAAGATAGAGACTTTGCCATGATATTAT	660
Qy	749	AAAAGCACTATGACCAATTTGSAAGCAATTTGTGCTATGAAAGAAAGAAAGCAATC	808
Db	661	AAAAGCACTATGACCAATTTGSAAGCAATTTGTGCTATGAAAGAAAGAAAGCAATC	720
Qy	809	AGTTGATGAATACCACTTTGTCTGAGAGCCCTTCAAGTGTGATCTGATCTGCCA	868
Db	721	AGTTGATGAATACCACTTTGTCTGAGAGCCCTTCAAGTGTGATCTGATCTGCCA	780
Qy	869	GGCATTTGAGATGCTTTACACACACCCTGTAATCTGCTGACAGTACCAACAGACGTG	928
Db	781	GGCATTTGAGATGCTTTACACACACCCTGTAATCTGCTGACAGTACCAACAGACGTG	840
Qy	929	GATTAATGTTCCATCACACAGAAAACTAATAGATTTCTCTGGAAATACAGCTGATATT	988
Db	841	GATTAATGTTCCATCACACAGAAAACTAATAGATTTCTCTGGAAATACAGCTGATATT	900
Qy	989	GCTACATCGTTCATCTGAGATGTTTGAAGTAAAGTATGACTTTTCAAAAGCTTTAA	1044
Db	901	GCTACATCGTTCATCTGAGATGTTTGAAGTAAAGTATGACTTTTCAAAAGCTTTAA	960
Qy	1049	ATTGTTAGAACATATCTACTAACTAAATTTCTGCTGACATATCCAAATATACAGAA	1110
Db	961	ATTGTTAGAACATATCTACTAACTAAATTTCTGCTGACATATCCAAATATACAGAA	1022
Qy	1109	TGTTATTCATCTAAAGCAATTTTTCATATCTCACTAAGATTAATTTAGCAATGCTTAA	1166
Db	1021	TGTTATTCATCTAAAGCAATTTTTCATATCTCACTAAGATTAATTTAGCAATGCTTAA	1088
Qy	1169	ATATCAAGAGTGTTCATTTGGAGTACCTGTAAATAGATGTGCAAGGGGACACATA	1222

Db	74531	ACGTTGCTACAAAGACCCCTTAAACCCCATTAACCAAGATGTGTAAGAAAAAGAAACCTTC	74590
QY	324	GCTATGTTGGCAATTTGTTCCCGTATTAAGATATATCTGGAACTATGAGGCCATCCCTC	383
Db	74591	GCTATGTTGGCAAAATTTGTTCCCGTATTAAGAGATATGTCGTGAACATTAAGTGCCATCTCC	74650
QY	384	AGACTTGGGAAGACCACCAAGGGGACACATATTAACATTAAGTGGGCTGTTGTGAGCAATGACC	443
Db	74651	AGACTTGGGAAGACCACCAAGGGGACACATATTAACATTAAGTGGGCTGTTGTGAGCAATGACC	74710
QY	444	CAATTGATGTGTGMAATTTGGAAGCAAGGTATGTGCAAGAGGTGGAATAATTTGGCTGA	503
Db	74711	CAATTGATGTGTGMAATTTGGAAGCAAGGTATGTGCAAGAGGTGGAATAATTTGGCTGA	74770
QY	504	AAAGTTCTAGGCATATTTGGCTATGATTTGACCAAGGGGAAACCCAGCTGGAAATGCTATGCA	563
Db	74771	AAAGTTCTAGGCATATTTGGCTATGATTTGACCAAGGGGAAACCCAGCTGGAAATGCTATGCA	74830
QY	564	TTAATGTGATGATGATCCTATGACAGCCAAATTTAAATGATATCAATGATGCTCAACGCGCTGA	623
Db	74831	TTAATGTGATGATGATCCTATGACAGCCAAATTTAAATGATATCAATGATGCTCAACGCGCTGA	74890
QY	624	AACCTGCGTACTTAGAAGCTACTGTGAGCTGCGTTTAGAAGGTATPAAGTTCCTGATGAA	683
Db	74891	AACCGGCTACTTAGAAGCTACTGTGAGCTGCGTTTAGAAGGTATPAAGTTCCTGATGAA	74950
QY	684	AACCGAATAAGTACGTTTGGCTTATATGAGATTTAAAGTATAGAGACTTTCCTCATGTATA	743
Db	74951	AACCGAATAAGTACGTTTGGCTTATATGAGATTTAAAGTATAGAGACTTTCCTCATGTATA	75009
QY	744	TTATTAAAGCACTCATGACCACTTGGAAAGCAATTAGTACTTAAGAAACGAATGGAAAG	803
Db	75010	TTATTAAAGCACTCATGACCACTTGGAAAGCAATTAGTACTTAAGAAAGTATGGAAAG	75069
QY	804	GAATAGTGTGATGAAATCAATTTTGCTGAGAGCCCTCAAGTGTGATCTGATGCTG	863
Db	75070	GAATAGTGTGATGAAATCAATTTTGCTGAGAGCCCTCAAGTGTGATCTGATGCTG	75129
QY	864	CCAGAGCCATTTGTGATGCTTTACCAACCAACCCTGTGAATTCGCTGCACAGTACCAACAG	923
Db	75130	CCAGAGCCATTTGTGATGCTTTACCAACCAACCCTGTGAATTCGCTGCACAGTACCAACAG	75189
QY	924	ACGTGATATAGTGTTCATTCACACGAAAAAAGTAATGATGATTTCTCTGGAAATACAAAGCTG	983
Db	75190	ACGTGATATAGTGTTCATTCACACGAAAAAAGTAATGATGATTTCTCTGGAAATACAAAGCTG	75249
QY	984	ATATTGCTACATCGTGTTTCATCTGCATGTATTAAGTAAGTAAGTATGAGCTTTTCAAGC	1043
Db	75250	ATATTGCTACATCGTGTTTCATCTGCATGTATTAAGTAAGTAAGTATGAGCTTTTCAAGC	75308
QY	1044	TTTAAATTTGTAGAACTCATCTTAAGTAAGTAATTTCTGCTGTGACTAATCCAAATATCT	1103
Db	75309	TTTAAATTTGTAGAACTCATCT---AGTAAATTTCTGCTGTGACCAATCCAAATATCT	75363
QY	1104	CAGATGTATATCATCATTAAGCAATTTTATATCTCAACATTAAGATTAACCTTTAGACACATG	1163
Db	75364	CAGATGTATATCATCATTAAGCAATTTTATATCTCAACATTAAGATTAACCTTTAGACACATG	75423
QY	1164	CTTAAATATCAAAAGAGTTGTCAATTTTGGAAAGTCACTTGTGAATAGATGTGCAAGGGGAGC	1223
Db	75424	CTTAAATATCAAAAGAGTTGTCAATTTTGGAAAGTCACTTGTGAATAGATGTGCAAGGGGAGC	75483
QY	1224	ACATATTGATGTATATGTTATACATATGTTAGGAATTAATAATTTTCT	1274
Db	75484	ACATATTGATGTATATGTTATACATATGTTAGGAATTAATAATTTTCT	75534

ACCESSION AC021159.4 GI:8572521
 VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 213025)
 TITLE Waterston, R.H.
 JOURNAL The sequence of Homo sapiens clone
 2 (bases 1 to 213025)
 REFERENCE Waterston, R.H.
 AUTHORS Direct Submission
 TITLE Submitted (14-JAN-2000) Genome Sequencing Center, Washington
 JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 COMMENT On Jun 19, 2000 this sequence version replaced gi:8570288.

----- Genome Center -----
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 41 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1384: contig of 1384 bp in length
 1385 1484: gap of unknown length
 1485 3382: contig of 1898 bp in length
 3383 3482: gap of unknown length
 3483 5013: contig of 1531 bp in length
 5014 5113: gap of unknown length
 5114 6422: contig of 1209 bp in length
 6423 6423: gap of unknown length
 6423 7736: contig of 1314 bp in length
 7737 7836: gap of unknown length
 7837 9198: contig of 1362 bp in length
 9199 9298: gap of unknown length
 9299 11720: contig of 2422 bp in length
 11721 11820: gap of unknown length
 11821 13496: contig of 1678 bp in length
 13497 13598: gap of unknown length
 13599 15122: contig of 1524 bp in length
 15123 15223: gap of unknown length
 15224 17332: contig of 2110 bp in length
 17333 17432: gap of unknown length
 17433 19823: contig of 2391 bp in length
 19824 19923: gap of unknown length
 19924 21811: contig of 1888 bp in length
 21812 21911: gap of unknown length
 21912 24564: contig of 2653 bp in length
 24565 24664: gap of unknown length
 24665 27708: contig of 3044 bp in length
 27709 27808: gap of unknown length
 27809 30176: contig of 2368 bp in length
 30177 30276: gap of unknown length
 30277 33515: contig of 3239 bp in length
 33516 33615: gap of unknown length
 33616 37205: contig of 3590 bp in length
 37206 37305: gap of unknown length
 37306 39846: contig of 2541 bp in length
 39847 39946: gap of unknown length
 39947 42384: contig of 2438 bp in length
 42385 42484: gap of unknown length
 42485 46260: contig of 3776 bp in length
 46261 46360: gap of unknown length
 46361 49107: contig of 2747 bp in length
 49108 49207: gap of unknown length
 49208 54501: contig of 5294 bp in length
 54502 54601: gap of unknown length
 54602 59315: contig of 4714 bp in length

59316 59415: gap of unknown length
 59416 65627: contig of 6212 bp in length
 65628 65727: gap of unknown length
 65728 72486: contig of 6758 bp in length
 72487 72586: gap of unknown length
 72587 77910: contig of 5325 bp in length
 77911 78010: gap of unknown length
 78011 84348: contig of 6338 bp in length
 84349 84448: gap of unknown length
 84449 90437: contig of 5989 bp in length
 90438 90537: gap of unknown length
 90538 97928: contig of 7391 bp in length
 97929 98028: gap of unknown length
 98029 105065: contig of 7037 bp in length
 105066 105165: gap of unknown length
 105166 111323: contig of 6158 bp in length
 111324 111423: gap of unknown length
 111424 117734: contig of 6311 bp in length
 117735 117834: gap of unknown length
 117835 125427: contig of 7593 bp in length
 125428 125527: gap of unknown length
 125528 134991: contig of 9464 bp in length
 134992 135091: gap of unknown length
 135092 143447: contig of 8356 bp in length
 143448 143547: gap of unknown length
 143548 156094: contig of 12547 bp in length
 156095 156194: gap of unknown length
 156195 157939: contig of 1745 bp in length
 157940 158039: gap of unknown length
 158040 175011: contig of 16972 bp in length
 175012 175111: gap of unknown length
 175112 187724: contig of 12613 bp in length
 187725 187825: gap of unknown length
 187826 208210: contig of 20366 bp in length
 208211 208311: gap of unknown length
 208312 213025: contig of 4715 bp in length.

FEATURES

source

misc_feature
 1
 /note="assembly_name:Contig11"
 misc_feature
 1485..3382
 /note="assembly_name:Contig14"
 misc_feature
 3483..5013
 /note="assembly_name:Contig15"
 misc_feature
 5114..6422
 /note="assembly_name:Contig16"
 misc_feature
 6423..7736
 /note="assembly_name:Contig18"
 misc_feature
 7837..9198
 /note="assembly_name:Contig24"
 misc_feature
 9299..11720
 /note="assembly_name:Contig25"
 misc_feature
 11821..13496
 /note="assembly_name:Contig26"
 misc_feature
 13599..15122
 /note="assembly_name:Contig27"
 misc_feature
 15223..17332
 /note="assembly_name:Contig28"
 misc_feature
 17433..19823
 /note="assembly_name:Contig30"
 misc_feature
 19924..21811
 /note="assembly_name:Contig31"
 misc_feature
 21912..24564
 /note="assembly_name:Contig32"
 misc_feature
 24665..27708
 /note="assembly_name:Contig33"
 misc_feature
 27809..30176
 /note="assembly_name:Contig34"
 misc_feature
 30277..59315

Db	75813	ACGTTGCTACAAAGACCCTTTAAACCCATTAAACAAGATGTGAAAAAAGAAACCTTC	75872
OY	324	GCTATGTTGGCAATTTGTTCCCGTATATAAGGATATATCTGGAACCTATGTGTCATCCCTC	383
Db	75873	GCTATGTTGGCAATTTGTTCCCGTATATAAGGATATATGTCGTGAACCTATGTGTCATCCCTC	75932
OY	384	AGACTGGGAAGACCCGAGGGCAGCATATATAACAATACTGCTGTGTCGTCGAACAATGACC	443
Db	75933	AGACTGGGAAGACCCGAGGGCAGCATATATAACAATACTGCTGTGTCGTCGAACAATGACC	75992
OY	444	CAATTTGATGTGTGTAATTTGGAAGCAAGGTATGTGCAGAGGTGAATATTTGGCTGA	503
Db	75993	CAATTTGATGTGTGTAATTTGGAAGCAAGGTATGTGCAGAGGTGAATATTTGGCTGA	76052
OY	504	AAGTTTACGATATTTGGCTATATGATTTACGAAGGGGAAACCGGACTGGAATGTCATGCCA	563
Db	76053	AAGTTTACGATATTTGGCTATATGATTTACGAAGGGGAAACCGGACTGGAATGTCATGCCA	76112
OY	564	TTAATGTGATGATCTCTGATGACACCAATTTATATGATATCAATGATGTCAAAAGGCTGA	623
Db	76113	TTAATGTGATGATCTCTGATGACACCAATTTATATGATATCAATGATGTCAAAAGGCTGA	76172
OY	624	AACCTGGCTACTTGAAGACTACTGTGAGCTGTGTTAGAAGATATAAGTGTCTGTATGAA	683
Db	76173	AACCGGCTACTTGAAGGCTACTGTGAGCTGTGTTAGAAGATATAAGTGTCTGTATGAA	76232
OY	684	AACCAAAAAAGACTTTGGCTTATATCCAGATTTAAAGATTAAGAGCTTGGCATTTGATA	743
Db	76233	AACCAAAAAAGACTTTGGCTTATATCCAGATTTAAAGATTAAGAGCTTGGCATTTGATA	76291
OY	744	TTATTTAAAAAGCAGCATGACCATTTGGAAAGCATTAGGACTTAAGAAAACGAATGAAAAAG	803
Db	76292	TTATTTAAAAAGCAGCATGACCATTTGGAAAGCATTAGGACTTAAGAAAATGAATGAAAAAG	76351
OY	804	GAATCATGTTGCATATATATACAACTTTGTCTGAGAGCCCTTCAGTGTGATCTCGATGCTG	863
Db	76352	GAATCATGTTGCATATACAAAACTGTGTCTGAGAGCCCTTCAGTGTGATCTCGATGCTG	76411
OY	864	CCAGAGCCATGTGTGATGCTTTTACACACACCCGTGTGATTTGGCTGTACAGATACCAACAG	923
Db	76412	CCAGAGCCATGTGTGATGCTTTTACACACACCCGTGTGATTTGGCTGTACAGATACCAACAG	76471
OY	924	ACGTGATTAAGTGGTTCCTATCACCAAAAACTAATAGATTTTCTGTGAATCAACGCTG	983
Db	76472	ACGTGATTAAGTGGTTCCTATCACCAAAAACTAATAGATTTTCTGTGAATCAACGCTG	76531
OY	984	ATATTGTACATCTGTTTCATCTGTGATGATTTAGAAATTAAGTAGAGCTTTTCAAGC	1043
Db	76532	ATATTGTGTGATCTGTTTCATCTGTGATGATTTAGAAATTAAGTAGAGCTTTTCAAGC	76590
OY	1044	TTTAAATTTGTAAGACTCATCTAATAAGTAAATTTGCTGTGACTAATCCATATACT	1103
Db	76591	TTTAAATTTGTAAGACTCATCTT-----AAGTAAATTTGCTGTGACTAATCCATATACT	76645
OY	1104	CAGATGTATTCACATTAAGACATTTTTCATATCTCAAGTAAGATTAAGTGTGACATG	1163
Db	76646	CAGATGTATTCACATTAAGACATTTTTCATATCTCAAGTAAGATTAAGTGTGACATG	76705
OY	1164	CTTAAATATCAACACAGTTGTCAATTTGGAAGTCACTTGTGAATAGATGTCGAAGGGAGC	1223
Db	76706	CTTAAATATCAACAGGAGTTGTCAATTTGGAAGTCACTTGTGAATAGATGTCGAAGGGAGC	76765
OY	1224	ACATATTTGATGATATATCTTACCATATGTTTNGAATAATTAATTTTGGCT	1274
Db	76766	ACATATTTGATGATATATCTTACCATATGTTTNGAATAATTAATTTTGGCT	76816
RESULT	12		
LOCUS	AC011012	155645 bp DNA HTG	03-NOV-2000
DEFINITION	Homo sapiens clone Rpl11-8D23, WORKING DRAFT SEQUENCE, 7 unordered pieces.		

Db 1022 TTTAAATTATAGACTAATC-----AGTCAGATTCTGTCGACCAATCCAGTGTATT 1075

Qy 1104 CAGATGTATCCATCTAAGCATTTTTCATATCTCACTAAGATTAACCTTTAGACATG 1163

Db 1076 CATACATTATCTATCTAAGACATTATTC--GTGCACTAAGGATAGACATTTAGTTCTG 1133

Qy 1164 CTTAAATATCAAGAGTGTTCATTTGAGAGTCTGTGATAGATGACAGGAGC 1223

Db 1134 CTTAAATATCAAGAGGCTGTGAGTTGAGTTGAGAGCATCTGTGACAGATGATTAAGT 1193

Qy 1224 ACATTTGATGTATATGTCTTACCATATGTTAGAAATTAATTTT 1271

Db 1194 ACATTTGATGTATATGTCTTACCATATGTTAGAAATTAATTTT 1241

RESULT 14

LOCUS AF108211 846 bp mRNA PRI 14-Apr-1999

DEFINITION Homo sapiens cytosolic inorganic pyrophosphatase mRNA, partial cds.

ACCESSION AF108211

VERSION AF108211.1 GI:4583152

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 846)

AUTHORS Rumsfeld,J., Ziegelbauer,K. and Spaltmann,F.

TITLE Cloning, expression, affinity purification and characterization of polyhistidine-tagged cytosolic Saccharomyces cerevisiae and human inorganic pyrophosphatases for differential screening of compounds for antifungal activity

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 846)

AUTHORS Rumsfeld,J., Ziegelbauer,K. and Spaltmann,F.

TITLE Direct Submission

JOURNAL Submitted (20-NOV-1998) Research Antinfectives 1, Bayer AG, P.O. Box 10179, Wuppertal 42096, Germany

FEATURES

location/Qualifiers

1. 846

/organism="Homo sapiens"

/db_xref="taxon:9606"

/cell_line="BJAB"

/cell_type="B-cell"

<1..>846

/note="iPpase"

/codon_start=1

/product="cytosolic inorganic pyrophosphatase"

/protein_id="AAD24964.1"

/db_xref="GI:4583153"

/translation="STEEARAASLEVRVPLKNKGOYISPEHDIPYADKDFHNV
EVRPNSNAKMEIATKDPINPIKODYKKKLYANLPPYKGTINNYGATPOTWEDPGH
NDKHTGCGGDDPIDVCEIGSVKARCEIIGVKVGLIAMIDEGETMKVLAINDVDS
DAANYNDINDVRLKPGYLEAVTFMFRYKVPDKPEKEFEAFKDKFAIDILKS
THDKMKAIVTKTKNGKISCMNTLTSESFPKCDPDAARAIVDALPPCESACTVPTDV
DKWFHH"

CDS

BASE COUNT 270 a 155 c 201 g 220 t

ORIGIN

Query Match 66.0%; Score 841.2; DB 9; Length 846;

Best Local Similarity 99.6%; Pred. No. 2e-184;

Matches 843; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 121 CACATGTAGTGAAGTACACAGCTGTGCTATATGCAAAAATGACATTCCTCAAGAGC 180

Qy 281 CCTTAAACCTTATTAACACAGATGTGAAAAAGAAAACCTTGCTATGTTCCGAATTTG 340

Db 181 CCTTAAACCTTATTAACACAGATGTGAAAAAGAAAACCTTGCTATGTTCCGAATTTG 240

Qy 341 TTCCCGTATTAAGATATATCTGGAAGTATGTTGCTCCATCCCTCAGACTTGGGAACCCA 400

Db 241 TTCCCGTATTAAGATATATCTGGAAGTATGTTGCTCCATCCCTCAGACTTGGGAACCCA 300

Qy 401 GGGCAATGATTAACATACACTGCTGTGTGTGTGCAATGACCAATGATGTGTGTA 460

Db 301 GGGCAATGATTAACATACACTGCTGTGTGTGTGCAATGACCAATGATGTGTGTA 360

Qy 461 ATTGAAGCAAGTATGTGCAAGAGTGAATAATTTGGCTGAAGTTCTAGGCATTTG 520

Db 361 ATTGAAGCAAGTATGTGCAAGAGTGAATAATTTGGCTGAAGTTCTAGGCATTTG 420

Qy 521 GCTATGATTTGACGAAGGGGAACCCAGCTGGAAGTCAATGCTTAATGTGATGATCT 580

Db 421 GCTATGATTTGACGAAGGGGAACCCAGCTGGAAGTCAATGCTTAATGTGATGATCT 480

Qy 581 GATGAGCCCAATTATTAATGATATCAATGATGTCAACAGGCTGAACCTGCTACTAGAA 640

Db 481 GATGAGCCCAATTATTAATGATATCAATGATGTCAACAGGCTGAACCTGCTACTAGAA 540

Qy 641 GCTACTGTGACTGCTTTGAAGAGTATTAAGTCTGATGAGAAAACAGAAAATGACTTT 700

Db 541 GCTACTGTGACTGCTTTGAAGAGTATTAAGTCTGATGAGAAAACAGAAAATGACTTT 600

Qy 701 GCGTTAATGCAAGATTTAAGATTAAGAGCTTGGCATGATATTAAGCACTCAT 760

Db 601 GCGTTAATGCAAGATTTAAGATTAAGAGCTTGGCATGATATTAAGCACTCAT 660

Qy 761 GACCATTTGGAAGCATTAAGTACTAAGAAAAGCAATGGAAGAAATGAGTTCATGAT 820

Db 661 GACCATTTGGAAGCATTAAGTACTAAGAAAAGCAATGGAAGAAATGAGTTCATGAT 720

Qy 821 ACACTTTGTCTGAGAGCCCTTCAAGTGTATCTGATGCTGCCAGACCATTTGGCAT 880

Db 721 ACACTTTGTCTGAGAGCCCTTCAAGTGTATCTGATGCTGCCAGACCATTTGGCAT 780

Qy 881 GCTTACCAACCACTGTGATCTGCTGACAGTACCAACAGAGCTGATTAAGTGTTC 940

Db 781 GCTTACCAACCACTGTGATCTGCTGACAGTACCAACAGAGCTGATTAAGTGTTC 840

Qy 941 CATCAC 946

Db 841 CATCAC 846

RESULT 15

LOCUS BC010468 1243 bp mRNA ROD 12-JUL-2001

DEFINITION Mus musculus, RIKEN cDNA 2010317E03 gene, clone MGC:6716

ACCESSION BC010468

VERSION BC010468.1 GI:14714656

KEYWORDS

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 1243)

AUTHORS Strausberg,R.

TITLE Direct Submission

JOURNAL Submitted (10-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Query Match	Similarity	65.6%	Score 835.8	DB 10	Length 1243
Best Local	Similarity	82.7%	Pred. No. 3.6e-183		
Matches 1020	Conservative	0%	Mismatches 197	Indels	Gaps
50	GTGGCTCTGTGGCAGCGCGCGCGGACGAGACTCCGGCATATGAGCGGGCTTCACACCGAG	109			
3	GTGGCTGTGCAGCGGGCGCGGGGTCACTCCGACACCATGAGCGGGCTTCACACCGAG	62			
110	GAGCGCGCGCGCGCGCTTCTCCTCGAGATACCGAGTTCCTCCAAATGAGAAAGGACAA	169			
63	GAGCGCGCGCGCGCTTCACCGCTGAGTACCGGGGTCTCTCTCAAAATGAGAAAGGCGAG	122			
170	TATATATCTCATTCATCATATATATCCAAATTATGAGATPAAGGATGTGTTTCACATGTA	229			
123	TACATTTCCTCATTTTATATATATGTTCCAAATTATGAGACAGAGATGTGTTCCACATGATG	182			
230	GTTGAAGTACCGACCGCTGCTTAATGACAAAATGAGATGCTCTCAAAAGACCCCTTTAAAC	289			
183	GTTGAGGTTTCCAGCGTGGTCCACGCGCCAAATATGAGATGTGCTCAAAAGACCCCTTTAAAC	242			
290	CCTATTAAACAGATGTGGAAGAAAGGAAACCTCGCATGTGCGAATTTGTTCCCGTAT	349			
243	CCATATCAAGCAAGATGTGAAAAAGGGGGAACCTCCGCTATGTGGGAATCTGTTCCTTAT	302			
350	AAAGGATATATCTGGAACATATGGAGCCATCCCTCAGACTGTGGGAAGACCCAGGCGACAT	409			
303	AAAGGATATATTTTGAACATACGGCGCCATCCCTCAGACATGGGAAAGCCAGAGACACAGT	362			

Search completed: March 29, 2002, 19:30:55
Job time: 2298 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2002, 18:53:47 ; Search time 90.12 Seconds
(Without alignments)
3204.166 Million cell updates/sec

Title: US-09-415-540-2
Perfect score: 1275
Sequence: 1 CAAGAGGTTGGGCTCTCT.....GAATATAATATTGCTG 1275

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1274	99.9	1275	2	US-08-741-437-2
2	1274	99.9	1275	2	US-09-134-593-2
3	51	4.0	7218	1	US-08-232-463-14
4	39.2	3.1	531	2	US-08-809-267-9
5	39.2	3.1	531	5	PCT-US95-13662A-9
6	37.6	2.9	1155	3	US-08-755-587-35
7	36.8	2.9	19124	2	US-08-487-826B-13
8	36.2	2.8	4507	2	US-08-568-459A-3
9	36.2	2.8	2781	3	US-08-487-826B-3
10	35.8	2.8	2726	1	US-08-749-522-4
11	35.8	2.8	2726	1	US-08-281-714-1
12	35.8	2.7	5555	1	US-08-484-338-3
13	34.8	2.7	9636	1	US-08-323-170B-1
14	34.8	2.7	29604	3	US-08-781-891-207
15	34.6	2.7	1939	1	US-07-715-751B-2
16	34.4	2.7	2186	2	US-08-878-546-9
17	34.4	2.7	11958	4	US-09-134-246-8
18	34.2	2.7	3022	3	US-08-961-083-215
19	34.2	2.7	3618	1	US-07-872-678A-36
20	34	2.7	756	3	US-08-513-974B-53
21	34	2.7	810	3	US-08-513-974B-53
22	34	2.7	1882	1	US-08-257-073-12
23	34	2.7	1884	1	US-08-257-073-8
24	34	2.7	2850	2	US-08-224-482-7
25	33.8	2.7	4403765	4	US-09-103-840A-2
26	33.6	2.6	950	3	US-08-581-148C-10
27	33.4	2.6	377	2	US-08-332-766A-1

C 28	33.2	2.6	1973	6	5256643-1	Patent No. 5256643
C 29	33.2	2.6	2675	1	US-07-749-001-4	Sequence 4, Appl
C 30	33.2	2.6	2675	1	US-08-154-198-4	Sequence 4, Appl
C 31	33.2	2.6	2675	1	US-08-463-335-4	Sequence 4, Appl
C 32	33.2	2.6	2675	2	US-08-464-023A-4	Sequence 4, Appl
C 33	33.2	2.6	5761	1	US-07-749-001-2	Sequence 2, Appl
C 34	33.2	2.6	5761	1	US-08-154-198-2	Sequence 2, Appl
C 35	33.2	2.6	5761	1	US-08-463-335-2	Sequence 2, Appl
C 36	33.2	2.6	5761	2	US-08-464-023A-2	Sequence 2, Appl
C 37	33.2	2.6	36741	4	US-09-301-665-3	Sequence 3, Appl
C 38	33	2.6	2573	3	US-08-714-918-17	Sequence 17, Appl
C 39	33	2.6	2573	3	US-08-714-918-64	Sequence 64, Appl
C 40	33	2.6	2573	4	US-09-265-315-17	Sequence 17, Appl
C 41	33	2.6	2573	4	US-09-265-315-64	Sequence 64, Appl
C 42	33	2.6	2573	4	US-09-265-315-17	Sequence 17, Appl
C 43	33	2.6	2573	4	US-09-265-315-64	Sequence 64, Appl
C 44	33	2.6	2573	4	US-09-266-417-17	Sequence 17, Appl
C 45	33	2.6	2573	4	US-09-266-417-64	Sequence 64, Appl

ALIGNMENTS

```
RESULT 1
US-08-741-437-2
: Sequence 2, Application US/08741437
: Patent No. 5843665
GENERAL INFORMATION:
APPLICANT: Hawkins, Phillip R.
APPLICANT: Hillman, Jennifer L.
TITLE OF INVENTION: A NOVEL HUMAN PYROPHOSPHATASE
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304
COMPUTER-READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/741,437
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0148 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 855-0555
TELEFAX: (415) 845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1275 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY:
CLONE: Consensus
US-08-741-437-2

Query Match 99.9%; Score 1274; DB 2; Length 1275;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


[illegible]

QY 549 GGAAGTCATTCGATTAATGTCATGTCATGACGCAATTAATGATCAATG 608
Db 1343 RRR 1284
QY 609 ATGCAACGCGTGAACCTGGCTACTTAGAGCTAGTGGAGCTGTTAGAGTATA 668
Db 1283 RRR 1224
QY 669 AGGTTCTGATGGAACCAAGAAATGAGTTGCTTAATGCAATTAAGATAGG 728
Db 1223 RRR 1164
QY 729 ACTTGCAATGATATATTAAGCAATGATGCAATGGAAGCTTAGTACTAGA 788
Db 1163 RRR 1104
QY 789 AAGCATGGAAGAAAGATGTCATGATCAATCACTTG 829
Db 1103 RRR 1063

RESULT 4

US-08-809-267-9
Sequence 9, Application US/08809267
Patent No. 5861296
GENERAL INFORMATION:
APPLICANT: LENNOX, Tricia L.
APPLICANT: SLATKO, Barton E.
APPLICANT: SEARS, Lauren E.
TITLE OF INVENTION: PURIFIED THERMOSTABLE INORGANIC
TITLE OF INVENTION: PYROPHOSPHATASE OBTAINABLE FROM THERMOCOCCUS
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEW ENGLAND BIOLABS, INC.
STREET: 32 TOZER ROAD
CITY: BEVERLY
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01915
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/809,267
FILING DATE: 12-MAR-1997
CLASSIFICATION: 433
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13662
FILING DATE:
APPLICATION NUMBER: US 08/329,721
FILING DATE: 25-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAMS, GREGORY D.
REGISTRATION NUMBER: 30901
REFERENCE/DOCKET NUMBER: NEB-105-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 927-5054
TELEFAX: (508) 927-1705
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 531 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-809-267-9

Query Match 3.1%; Score 39.2; DB 2; Length 531;
Best Local Similarity 50.5%; Pred. No. 0.1;

Matches 95; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
QY 425 TGTGTGTCACATATGACCAATGTCGTGTGGAATTGGAAGCAAGTATGCAAGA 484
Db 187 TGTAGCATATGACACACCGTTTGACATCATGTCATATGATGAGCAACATATCCG 246
QY 485 GGGAATATATGCGGTGGAAGTTTACGATATGTCATGATGATGACGAGGGAAGG 544
Db 247 GGAGTCTTATTTAGGCAAGACCAATAGGCTTTCAGATGATGATGACAGCGGCAAG 306
QY 545 GACTGGAAGTCATTCATTAATGTCATGATGTCATGATGACGCCAATTAATGATATC 604
Db 307 GACTACAGGATATGTCAGTTCAGTGAAGATCCCTACTTAAATGACTGGAAGACATA 366
QY 605 AATGATGT 612
Db 367 AGCGACGT 374

RESULT 5

PCT-US95-13662A-9
Sequence 9, Application PC/TUS9513662A
GENERAL INFORMATION:
APPLICANT: LENNOX, Tricia L.
APPLICANT: SLATKO, Barton E.
APPLICANT: SEARS, Lauren E.
TITLE OF INVENTION: PURIFIED THERMOSTABLE INORGANIC
TITLE OF INVENTION: PYROPHOSPHATASE OBTAINABLE FROM THERMOCOCCUS
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEW ENGLAND BIOLABS, INC.
STREET: 32 TOZER ROAD
CITY: BEVERLY
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01915
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13662A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,721
FILING DATE: 25-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAMS, GREGORY D.
REGISTRATION NUMBER: 30901
REFERENCE/DOCKET NUMBER: NEB-105-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 927-5054
TELEFAX: (508) 927-1705
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 531 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
PCT-US95-13662A-9

Query Match 3.1%; Score 39.2; DB 5; Length 531;
Best Local Similarity 50.5%; Pred. No. 0.1;
Matches 95; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
QY 425 TGTGTGTCACATATGACCAATGTCGTGTGGAATTGGAAGCAAGTATGTCAGAA 484
Db 187 TGTAGCATATGACACACCGTTTGACATCATGTCATATGATGAGGAGCAACATATCCG 246

DB 5490 TTTTATGTCATTTATGTTATATATATTAACATGATGTTTTTGAANAAT 5549
OY 1072 ACTAATTCCTGCTGACATAATACACAAAGTTATCCATCAAGCATTTT 1131
DB 5550 ATTAAATATCATATATATATATATATATATATATATATATATATAT 5609
OY 1132 CATATCTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1191
DB 5610 TAACTAT 5669
OY 1192 AAGTACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1251
DB 5670 TGAATATAGGCTATGATGATGATGATGATGATGATGATGATGATGAT 5729
OY 1252 TTAGCAAT 1275
DB 5730 CATATAGTATATATATATATATATATATATATATATATATATATAT 5753

RESULT 8

US-08-568-459A-3

Sequence 3, Application US/08568459A
Patent No. 5849306

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4507 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum

US-08-568-459A-3

Query Match 2.8%; Score 36.2; DB 2; Length 4507;
Best Local Similarity 56.2%; Pred. No. 2.2;
Matches 68; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

OY 720 AAGTAAAGGACTTGCATGATATATTAAGACACTGACCATGGAAGCATTA 779
DB 2286 AAGTAAAGGATGACCAATTTCTATTAATGAAGAAATGAACCAACTTCC 2345
OY 780 TGACTAAGAAACGATGGAAGGATGATGATGATGATGATGATGATGATG 839
DB 2346 TTCTGAGGAAACGATGATGATGATGATGATGATGATGATGATGATG 2405
OY 840 C 840
DB 2406 C 2406

RESULT 9

US-08-487-826B-3

Sequence 3, Application US/08487826B
Patent No. 5998827

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4507 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum

US-08-487-826B-3

Query Match 2.8%; Score 36.2; DB 2; Length 4507;
Best Local Similarity 56.2%; Pred. No. 2.2;
Matches 68; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
OY 720 AAGTAAAGGACTTGCATGATATTAAGACACTGACCATGGAAGCATTA 779
DB 2286 AAGTAAAGGATGACCAATTTCTATTAATGAAGAAATGAACCAACTTCC 2345
OY 780 TGACTAAGAAACGATGGAAGGATGATGATGATGATGATGATGATGATG 839
DB 2346 TTCTGAGGAAACGATGATGATGATGATGATGATGATGATGATGATG 2405

APPLICANT: Hellstr m, Karl E.
TITLE OF INVENTION: HERA HUMAN RECEPTOR TYROSINE KINASE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,438
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/323,442
FILING DATE: 14-OCT-1994
APPLICATION NUMBER: US 08/150,704
FILING DATE: 10-NOV-1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/981,165
FILING DATE: 24-NOV-1992
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 5624-230
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5555 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 34..3210
US-08-484-438-3

Query Match 2.7%; Score 35; DB 1; Length 5555;
Best Local Similarity 49.8%; Pred. No. 5.3; Mismatches 115; Indels 2; Gaps 1;
Matches 116; Conservative 0;
DB 1040 AACCTTAATTTGTGACCTCATCTAACAAGTAATTCGCTGCTAATCAATA 1099
DB 3480 AATGGTTTATTGTTGGAATTTGTTTAAAGTTAAGTTTGTAAATAATTAAGA 3421
DB 1100 TACTGAAATGTATTCATCAATCAAGCAATTTTCATATCTCAACTAAGTAACCTTTTGA 1159
DB 3420 TGAGGAAATGTTCTCTGTAATCTTGAAGGTATTTCAAGAGTACATGTTGATTTGTC 3361
DB 1160 CATGCTTAATATCAAGCAAGTGTCTATTTGAGAGTCACTGTGATATGATGCAAGG 1219
DB 3360 CATGCTTGCTTCCAAATTAACCTACATATAGCA--AGATGGACATCTTTTGTGGGTA 3303
DB 1220 GAGCACATATGATGATATGTTACCATATGTTAGGAATAAATAATTTTGG 1272
DB 3302 TGGTATATTTTATCATGAGCATCACATGATTTGAGCATATATATATTTTGG 3250

RESULT 13
US-08-323-170B-1/c

Sequence 1, Application US/08323170B
Patent No. 573772
GENERAL INFORMATION:
APPLICANT: Williamson, Kim C.
ADDRESSEE: Kaslow, David C.
TITLE OF INVENTION: Cloning and Expression of Plasmidum
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,170B
FILING DATE: 13-OCT-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/010,409
FILING DATE: 29-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Quine, Jonathan A.
REGISTRATION NUMBER: P-41,261
REFERENCE/DOCKET NUMBER: 015280-113100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 149..9556
US-08-323-170B-1

Query Match 2.7%; Score 34.8; DB 1; Length 9636;
Best Local Similarity 48.1%; Pred. No. 8; Mismatches 107; Indels 0; Gaps 0;
Matches 99; Conservative 0;
DB 950 AAAAATTAATGAGATTCTCTGGAATACAGCTGATATGCTACATCGTTCATCTGGA 1009
DB 2773 AGAAATTAATTAATCTCTCTCTTTTAATATATATTAATTAATTAATTAATGATTC 2714
DB 1010 TGTATTAGAAATGAAAGTAGAGCTTTCAAGCTTTAAATTTGAGAGCACTCACTAAT 1069
DB 2713 TTTAATACCTTGAGAAATGCTGAGCACTGAGCAATTAATAATTTTGCATTCATCTCTT 2654
DB 1070 AAGTAATATCTGCTGCTGACATCAATCAATATCACTGAGATGTTATTCATCAAGCATTT 1129
DB 2653 ATTATATCTACAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2594
DB 1130 TTCATATCTCACTAAGTAATCTTTT 1155
DB 2593 TATTTCTTATTAATTAAGTAATCATCT 2568

RESULT 14
US-08-781-891-207/c
Sequence 207, Application US/08781891
Patent No. 6090620
GENERAL INFORMATION:

APPLICANT: Fu, Ying-Hui
 APPLICANT: Yu, Chang-Ea
 APPLICANT: Oshima, Junko
 APPLICANT: Mulligan, John T.
 APPLICANT: Schellenberg, Gerald D.
 TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
 TITLE OF INVENTION: WERNER'S SYNDROME
 NUMBER OF SEQUENCES: 209
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/781,891
 FILING DATE: 27-DEC-1996
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: No. 6090620tenburg Ph.D., Carol
 REGISTRATION NUMBER: 39,317
 REFERENCE/DOCKET NUMBER: 240052.419
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 207:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 29604 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2002, 18:53:17 ; Search time 1498.3 Seconds

(Without alignments)
9144.275 Million cell updates/sec

Title: US-09-415-540-2

Sequence: 1275
1 CAAGAGGTTCGGCTCTCT.....GAATATAATTTTTCGTCG 1275

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

EST:
1: em_estfun:
2: em_esthum:
3: em_estin:
4: em_estom:
5: em_estipl:
6: em_estda:
7: em_estro:
8: em_estov:
9: em_hic:
10: gb_estl:
11: gb_estc2:
12: gb_hic:
13: gb_gss:
14: em_gss_fun:
15: em_gss_hum:
16: em_gss_inv:
17: em_gss_pla:
18: em_gss_pro:
19: em_gss_rtd:
20: em_gss_vit:
21: em_gss_other:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	933.6	73.2	978	10	AL525341
2	898.8	70.5	985	10	AL574658
3	836.2	65.6	1293	12	AK008575
4	813	63.8	826	10	AL519324
5	811.6	63.7	932	11	BG676397
6	776.8	60.9	826	10	AL515056
7	760	59.6	996	11	BG674343
8	742.8	58.3	773	11	BG715195
9	742	58.2	771	11	BI258591
10	739.8	58.0	772	11	BG713976
11	732.8	57.5	878	11	BI259308
12	730.2	57.3	1308	12	AK020291

13	721	56.5	915	11	BG574468
14	720.4	56.5	845	11	BG708966
15	719.4	56.4	762	11	BG721093
16	712.6	55.9	804	11	BI260339
17	709	55.6	714	2	BG700549
18	707.2	55.5	754	11	BG699988
19	707.2	55.5	1076	11	BG179769
20	706.8	55.4	836	11	BG715251
21	706.2	55.4	950	11	BG704401
22	705.8	55.4	955	11	BF791373
23	705.4	55.3	805	11	BF967952
24	704.2	55.2	769	11	BG702317
25	703.6	55.2	751	10	AI078153
26	691.6	54.2	771	10	AI956149
27	688.8	54.0	745	10	AL550933
28	687	53.9	828	11	BG705302
29	675.4	53.0	711	11	BG777143
30	672.8	52.8	734	11	BE797649
31	672.2	52.7	998	11	BF966851
32	670.2	52.6	781	11	BG108210
33	667.4	52.3	714	11	BG702955
34	663.4	52.0	929	11	BF793034
35	662.2	51.9	775	11	BG036446
36	660	51.8	693	11	BG703149
37	659.6	51.7	847	11	BI093604
38	658.4	51.6	740	11	BG497509
39	656	51.5	843	11	BG106064
40	654.8	51.4	905	11	BG027821
41	653	51.2	997	11	BG255623
42	650.8	51.0	659	10	AI744359
43	648.8	50.9	676	11	BG499929
44	648	50.9	921	11	BG282637
45	645.8	50.7	677	10	AL519323

ALIGNMENTS

RESULT 1
AL525341
LOCUS
DEFINITION
AL525341 LTL_NFL003.NBC3 Homo sapiens CDNA clone CS0DC011YC02 5
prime, mRNA sequence.
ACCESSION
AL525341
VERSION
AL525341.1 GI:12788834
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 978)
L.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length CDNA libraries and normalization
Unpublished (2001)
JOURNALS
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 978

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DC011YC02"
/clone_id="LTL_NFL003.NBC3"
/sex="male"
/tissue-type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang life

Query Match	73.2%;	Score 933.6;	DB 10;	Length 978;
Best Local Similarity	97.7%;	Pred. No. 8.5e-218;		
Matches 956;	Conservative 10;	Mismatches 11;	Indels 2;	Gaps 2

D_b. 901 TTACAGCTGCTAAATATC-AACCAATTGCATTTGGAGTCACCTGTATAATAGATGTT 958
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
QY 1214 CAAGGGAGACATATTGG 1232
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
D_b: 960 CCAAGGGAGCCCATATTGG-978

```

RESULT      2
LOCUS       AL574658/c
DEFINITION  AL574658      985 bp      mRNA
VERSION     AL574658 L17_NFL006_PL2 Homo sapiens cDNA clone CS001065YG18.3
ACCESSION   AL574658
KEYWORDS    prime, mRNA sequence.
SOURCE      AL574658.1  GI:12935068
            EST.
            human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE   1 (bases 1 to 985)
AUTHORS     Li W.B., Gruber,C., Jesse,J. and Polyes,D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
            source
            1..985

```

FEATURES	SOURCE	LOCATION/QUALIFIERS
BASE COUNT	261 a	1. 985
ORIGIN	212 c 162 g 322 t 28 others	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="CSOD1065Y618" /clone_11b="LY1_NFL006_PL2" /tissue_type="Placenta" /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dt) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8571 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com "

	Query Match	70.5%	Score 98.8	DB 10	Length 985
	Best Local Similarity	96.6%	Pred. No. 2.8e-209		
	Matches 910	Conservative 26	Mismatches 4	Indels 2	Gaps 2
QY	281	CCTTTAAACCCATTTAAACACAGATGTGAAAAAGAAAACCTTCGTATGTTCGAATTTG	340		
DB	953	CCCTTAAACCCATTTAAACAGATGTGAAAAAGAAAAC-TCGCTATGTTCGCAATTTG	895		
QY	341	TTCCCGTATTAAGATATATCTGGAACATATGTCGCATCCCTCAGACTTGGAGACCCA	400		
DB	894	TTCCCGTATTAAGATATATCTGGAACATATGTCGCATCCCTCAGACTTGGAGACCCA	835		
QY	401	GGGCACATGATTAACATACCTGCTGTGTGGTGACAAATGA-CCCAATTGATGTGTGTA	459		
DB	834	GGGCACATGATTAACATACCTGCTGTGTGTGTGCATAATGACCCCAATTGATGTGTGTA	775		
QY	460	AATTGGAAGCAAGGTATGTGCAAGAGGTGAATAATTTGGCGTGAAGAGTTCTAGGCATATT	519		
DB	774	AATTGGAAGCAAGGTATGTGCAAGAGGTGAATAATTTGGCGTGAAGAGTTCTAGGCATATT	715		
QY	520	GGCATATGATTTACCAAGGGGAAACCGACTGGAAGTCATTTGCATTAATGTGGATGATCC	579		
DB	714	GGCATATGATTTACCAAGGGGAAACCGACTGGAAGATCTTTGCATTAATGTGGATGATCC	655		

QY	580	TGATCACCACCAATTTAATGATATCAACATGATGTCATAAAGCGGTGAACCTGGCTCTACTAGA	639
Db	654	TGATGACACCCATATATATGATATCAACAGATGTCATAAAGCGGTGAACCTGGCTCTACTAGA	595
QY	640	AGCTACTGTGACGTGTTTATAGAGGTATAGGTTCTTGATGAAAAACCCAGAAAATGAGTT	699
Db	594	AGCTACTGTGACGTGTTTATAGAGGTATAGGTTCTTGATGAAAAACCCAGAAAATGAGTT	535
QY	700	TGCGTTTAATCAGAAATTAAGAATTAAGGACTTTGGCATTTATTTAATTAACCACTCA	759
Db	534	TGCGTTTAATCAGAAATTAAGAATTAAGGACTTTGGCATTTATTTAATTAACCACTCA	475
QY	760	TGACCATTTGGAAGACATTTAGTACTAGAGAAAAACGAAATGGAAGAAATCACTTCATGAA	819
Db	474	TGACCATTTGGAAGACATTTAGTACTAGAGAAAAACGAAATGGAAGAAATCACTTCATGAA	415
QY	820	TACAACTTTGTCGTAGAGACCCCTTCAGTGTGATCTGTATCTGCGCAGAGCCATTTGGA	879
Db	414	TACAACTTTGTCGTAGAGACCCCTTCAGTGTGATCTGTATCTGCGCAGAGCCATTTGGA	355
QY	880	TGCTTTTACCACACCCCTTGATGATGCGCTGACAGTACCAACAGCGGATTAAGTGGTT	939
Db	354	TGCTTTTACCACACCCCTTGATGATGCGCTGACAGTACCAACAGCGGATTAAGTGGTT	295
QY	940	CCATCACCAGAAAACTAATGAGATTTTCTCTCGAATACAAACCTGATTTGCTACATGCTG	999
Db	294	CCATCACCAGAAAACTAATGAGATTTTCTCTCGAATACAAACCTGATTTGCTACATGCTG	235
QY	1000	TTTCATCTGTGATGATTTAGAGATTAAGATAGTACGCTTTTCAAAGCTTTAATTTGTAGAAC	1059
Db	234	TTTCATCTGTGATGATTTAGAGATTAAGATAGTACGCTTTTCAAAGCTTTAATTTGTAGAAC	175
QY	1060	TCATCTAATCAAGAAATTTGCTGTGTACTAATTCATTAATCTACAGAAATTTATTCATC	1119
Db	174	TCATCTAATCAAGAAATTTGCTGTGTACTAATTCATTAATCTACAGAAATTTATTCATC	115
QY	1120	TAAAGCATTTTTCATATCTCACTAGATTAATCTTTTACGACATGCTTAAATATCAAGCA	1179
Db	114	TAAAGCATTTTTCATATCTCACTAGATTAATCTTTTACGACATGCTTAAATATCAAGCA	55
QY	1180	GTTCATTTTGGAGTCACTTGTGATATGATGATGCAAGGCA	1221
Db	54	GTTCATTTTGGAGTCACTTGTGATATGATGATGCAAGGCA	13
RESULT	3		
LOCUS	AK008575	1293 bp	mRNA
DEFINITION			AK008575 Mus musculus adult male small intestine cDNA, RIKEN full-length enriched library, clone:2010317E03, full insert sequence.
ACCESSION	AK008575		
VERSION	AK008575.1	GI:12842842	
KEYWORDS			CAP trapper.
SOURCE			Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to mRNA, clone:lib.RIKEN full-length enriched mouse cDNA library clone:2010317E03.
ORGANISM			Mus musculus
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS			1 (bases 1 to 1293)
TITLE			Carninci, P. and Hayashizaki, Y.
JOURNAL			High-efficiency full-length cDNA cloning
MEDLINE			Methods in enzymology. 303, 19-44 (1999)
PUBMED			99279253
REFERENCE			10349636
AUTHORS			2 (bases 1 to 1293)
TITLE			Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
JOURNAL			Itch, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
MEDLINE			Normalization and subtraction of cap-trapper-selected cDNAs to
PUBMED			prepare full-length cDNA libraries for rapid discovery of new genes
			Genome research. 10 (10), 1617-1630 (2000)
			20499374
			11042159

REFERENCE	3 (bases 1 to 1293)
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagoka,S., Sasaki,N., Carlini,J., Konno,H., Akiyama,J., Nishi,K., Kikunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishi,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamamoto,R., Matsunoto,H., Sakauchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Itawa,M., Ohtsuka,E., Watanishi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuyama,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed sequencing
JOURNAL	Genome research. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCES	4 (bases 1 to 1293)
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	5 (bases 1 to 1293)
AUTHORS	Adachi,T., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arikawa,T., Carlini,J., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hirooka,T., Hori,F., Imotani,K., Ishi,Y., Itoh,M., Izawa,M., Kato,H., Kanai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toyota,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute: 1-7-22 Shohitcho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gscc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
FEATURES	Location/Qualifiers
SOURCE	1. 1293
	/organism="Mus musculus"
	/strain="C57BL/6J"
	/db_xref="taxon:10090"
	/db_xref="MGD:MGI:1898257"
	/db_xref="MGD:MGI:1915145"
	/clone="2010317E03"
	/sex="male"
	/tissue_type="small intestine"
	/clone_lib="RIKEN full-length enriched mouse cDNA library"
	/dev_stage="adult"
	109. 978
CDS	/note="putative"
	/codon_start=1
	/protein_id="BAB25754.1"
	/db_xref="GI:12842843"
	/translation="MSGFSSSEERAPFTLEAYRVLKNEKGOYISPFHDVPIYADKWEFHMVVEPRMSNAKKEIAIKDPLNLIKODVKGKIRVYANLFPGYITWNGAIQOTWEDPGHSGICGCCGNDPIDVEIGSKSAREEIIIRVVLGILAMIDGSEPMKVIAINVDPSDANYNKDISGVNRLKPYELATYDMFPRYVDPDGGKLPENEFAPNAEKNKQFAVDIIRKTHDWKLVTKTKTDGKICSCSMNTTVSESPKCPDPAKAIYDALPPCESACSLPTVDKMKHPOOKN"
BASE COUNT	354 a 288 c 336 g 315 t
ORIGIN	
Query Match	65.6% Score 836.2; DB 12; Length 1293;
Best Local Similarity	82.4% Pred. No. 6e-194;
Matches 1024; Conservative	0; Mismatches 203; Indels 16; Gaps 5;
40 GTGCGGGCTGTGTGTGTGGCAGCGCGCGAGAGACTTCGGCACTATGACGGGCTT	99

```

Db 60 GTGCTGTTGTGCTGTGACGCGCGCGGCTGATCTCCGACACCATGAGCGCTT 119
Oy 100 CAGCACCAGGAGCGCGCGCGCTTCTCCCTGAGTACGAGTATCCCTCAAAATGA 159
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 120 CAGCAGCGAGGAGCGCGCGCGCTTCTCCCTGAGTACGAGTATCCCTCAAAATGA 179
Oy 160 GAAGGACCAATATATCTCCATTTCCATGATGTTCCATTTATGACAGAGATGTGT 219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 180 GAAGGACCAATATATCTCCATTTCCATGATGTTCCATTTATGACAGAGATGTGT 239
Oy 220 TCACATGTAGTGAAGTACACGCTGCTATATGCAAAATGAGATGTGTACAAAGA 279
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 240 CACACATGTGTGAGTGTCCACGCTGTGTCACCCCAAAATGAGATGTGTACAAAGA 299
Oy 280 CCGTTAAACCCATTAACAGATGTGAAAAAGAAAACCTTCCCTATGTTGCAATTT 339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 300 CCGTTAAACCCATTAACAGATGTGAAAAAGAAAACCTTCCCTATGTTGCAATTT 359
Oy 340 GTTCCCGTATAAGATATATCTGAACTATGCTGCTCCCTGACACTTGGGAGAGCCG 399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 360 GTTCCCGTATAAGATATATCTGAACTATGCTGCTCCCTGACACTTGGGAGAGCCG 419
Oy 400 AGGACACATGATTAACATCTGCGTGTGTGTGACATGACCCCAATGATGTGTGA 459
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 420 AGGACACATGATTAACATCTGCGTGTGTGTGACATGACCCCAATGATGTGTGA 479
Oy 460 AATTGGAAGCAAGTATGTCAGAGGTGAAATATTTGGCTGAAAGTTCTAGCATATT 519
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 480 AATTGGAAGCAAGTATGTCAGAGGTGAAATATTTGGCTGAAAGTTCTAGCATATT 539
Oy 520 GCGTATGATGAGCAAGGAGGAAACCGACTGAAAGTCAATTAATGTGATGATCT 579
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 540 GCGTATGATGAGCAAGGAGGAAACCGACTGAAAGTCAATTAATGTGATGATCT 599
Oy 580 TGATCCAGCCCAATTAATGATATCAATGATGCAACAGGCTGAAACCGCTACTTGA 639
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 600 TGATCCAGCCCAATTAATGATATCAATGATGCAACAGGCTGAAACCGCTACTTGA 659
Oy 640 AGCTCTGTGAGCTGTTAGAGGTATAGGTTCTGATGAGAAACCGCTACTTGA 699
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 660 AGCTCTGTGAGCTGTTAGAGGTATAGGTTCTGATGAGAAACCGCTACTTGA 719
Oy 700 TCCGTTTATGAGATTTAAAGATAGAGCTTTGCCATGATATTTTAAACCACTCA 759
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 720 TCCGTTTATGAGATTTAAAGATAGAGCTTTGCCATGATATTTTAAACCACTCA 779
Oy 760 TGACCTTGAAGATTAAGTACTAAGAAAACGAAATGAGAAAGAAATCACTTGAA 819
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 780 TGACCTTGAAGATTAAGTACTAAGAAAACGAAATGAGAAAGAAATCACTTGAA 839
Oy 820 TACACCTTGTGTGAGAGCCCTTCAAGTGTATCTGATGCTGCAAGGCACTTGTGA 879
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 840 CACACAGGCTGTGAGAGCCCTTCAAGTGTATCTGATGCTGCAAGGCACTTGTGA 899
Oy 880 TGGTTTACACACCGCTGTGAGATCTGCTGCAAGTGTATCTGATGCTGCAAGGCA 939
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 900 TGGTTTACACACCGCTGTGAGATCTGCTGCAAGTGTATCTGATGCTGCAAGGCA 959
Oy 940 CCATCACCAGAAAACCTATGAGATTTCTGGAATGAAAGGATATGCTACATCTGT 999
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 960 CCATCACCAGAAAACCTATGAGATTTCTGGAATGAAAGGATATGCTACATCTGT 1019
Oy 1000 TTTCTGAGTATGATTAAGATTAAGATTTTCAAGCTTTTCAAGCTTTTGAAGAG 1059
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1020 TTTCTGAGTATGATTAAGATTAAGATTTTCAAGCTTTTCAAGCTTTTGAAGAG 1073
Oy 1060 TCATCTAATTAAGTAAATCTGCTGATGATTAATTAATTAATTAATTAATTAAT 1119
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1074 TCATCTAATTAAGTAAATCTGCTGATGATTAATTAATTAATTAATTAATTAAT 1132
Oy 1120 TAAAGCATTTT-----TCATCTAATTAAGTAAATTAATTAATTAATTAATTAAT 1172
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1133 GTTACCATCTAGACACTGATTAATTAATTAATTAATTAATTAATTAATTAAT 1191

```

```

Oy 1173 CAAGACCATGTCATTTGAGTCACTTGTGAATAGATGTGCAAGGAGGACATATGG 1232
Db 1192 CAAGCCGTGGGAGACATGAGATGATGATGATGATGATGATGATGATGATGATG 1251
Oy 1233 ATGATATGTTACATATGATGATGATGATGATGATGATGATGATGATGATGATG 1275
Db 1252 ATGATATGTTACATATGATGATGATGATGATGATGATGATGATGATGATGATG 1293

RESULT 4
AL519324
LOCUS AL519324 LTL_NF011_NBC1 Homo sapiens cDNA clone CS0DA012YD02 5
DEFINITION prime, mRNA sequence.
ACCESSION AL519324
VERSION AL519324.1 GI:12782817
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 826)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
Location/Qualifiers
FEATURES
source
1. 826
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DA012YD02"
/clone_lib="LTL_NF011_NBC1"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a Notti-oligo(dT) primer. Five prime end
was enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library is not normalized, but is the control for
the normalized libraries. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
f.liang@lifestech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 257 a 153 c 203 g 212 t 1 others
ORIGIN

Query Match 63.8%; Score 813; DB 10; Length 826;
Best local similarity 99.6%; Pred. No. 2.5e-188;
Matches 824; Conservative 1; Mismatches 1; Indels 1; Gaps 1;
Oy 47 CTGCTGCTCTGTCGAGCGGCGGCGGAGCACTCCGCACTATGAGCGCTTCAGACCC 106
Db 1 CTGCTGCTCTGTCGAGCGGCGGCGGAGCACTCCGCACTATGAGCGCTTCAGACCC 60
Oy 107 GAGAGCGCGCGCGCGCTTCTCCCTGAGTACCGGATTTCTCTCAAAAATGAGAAAGA 166
Db 1 GAGAGCGCGCGCGCGCTTCTCCCTGAGTACCGGATTTCTCTCAAAAATGAGAAAGA 119
Oy 61 GAGAGCGCGCGCGCGCTTCTCCCTGAGTACCGGATTTCTCTCAAAAATGAGAAAGA 119
Db 167 CAATATATATCTCATTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 226
Db 120 CAATATATATCTCATTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 179
Oy 227 GTAGTTAAGTACACCGCTGTCTAATGCAAAAATGAGATTTCTCAAGAGACCCCTTAA 286
Db 180 GTAGTTAAGTACACCGCTGTCTAATGCAAAAATGAGATTTCTCAAGAGACCCCTTAA 239

```

```

QY 287 AACCTATTAAACAGATGTGAAAAAGAAACCTTGTGATGTCGCAATTTGTCGG 346
Db 240 AACCTATTAAACAGATGTGAAAAAGAAACCTTGTGATGTCGCAATTTGTCGG 299
QY 347 TATTAAGGATATATCTGGAACATATGTCCTCCCTCAGACTTGGGAAGACCCAGGCAC 406
Db 300 TATTAAGGATATATCTGGAACATATGTCCTCCCTCAGACTTGGGAAGACCCAGGCAC 359
QY 407 AATGATTAACATCTGCTGCTTGTGCTGACAAATGACCCATTTGATGCTGTAATTTGA 466
Db 360 AATGATTAACATCTGCTGCTTGTGCTGACAAATGACCCATTTGATGCTGTAATTTGA 419
QY 467 AGCAAGTATGTCGAAGAGTGAATATATGCGCTGAAGTTCAGACATATTTGCTATG 526
Db 420 AGCAAGTATGTCGAAGAGTGAATATATGCGCTGAAGTTCAGACATATTTGCTATG 479
QY 527 ATTGACGAAGGGGAACCGACTGGAAGTCAATTCCTTAATGTGATGATCCTGATGCA 586
Db 480 ATTGACGAAGGGGAACCGACTGGAAGTCAATTCCTTAATGTGATGATCCTGATGCA 539
QY 587 GCCAATTATATGATATCATGATGATGCAAAAGCGCTGAACCTGCTACTTGAAGTACT 646
Db 540 GCCAATTATATGATATCATGATGATGCAAAAGCGCTGAACCTGCTACTTGAAGTACT 599
QY 647 GTGCACTGCTTTGAAGAGTATTAAGCTTCTGATGGAAGAACAGAAATGATTTGCTTT 706
Db 600 GTGCACTGCTTTGAAGAGTATTAAGCTTCTGATGGAAGAACAGAAATGATTTGCTTT 659
QY 707 AATGACGAATTTTAAGATTAAGGACTTTGCCATGATATTTTAAGACACTCATGACAT 766
Db 660 AATGACGAATTTTAAGATTAAGGACTTTGCCATGATATTTTAAGACACTCATGACAT 719
QY 767 TGGAAACATTAGTATGATTAAGAAAGCATGGAAGAAAGATTCATGATGATGACAT 826
Db 720 TGGAAACATTAGTATGATTAAGAAAGCATGGAAGAAAGATTCATGATGATGACAT 779
QY 827 TTGTCTGAGAGCCCTTCAAGTGTGATCTGATGCTGCCAGAGCCAT 873
Db 780 TTGTCTGAGAGCCCTTCAAGTGTGATCTGATGCTGCCAGAGCCAT 826

```

```

RESULT 5
LOCUS BG676397 932 bp mRNA EST 01-MAY-2001
DEFINITION 602622839F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:474797 5',
ACCESSION BG676397
VERSION BG676397.1 GI:13907793
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 932)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
http://image.llnl.gov
Plate: L1M10598 row: n column: 10
High quality sequence stop: 837.
location/Qualifiers
1..932
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

FEATURES

```

source
1..932
/db_xref="taxon:9606"

```

```

/clone="IMAGE:474797"
/clone_id="NCI_CGAP_Skn4"
/tissue="squamous cell carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Origin: skin; Vector: pCMV-Sport6; Site: 1: Not;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 309 a 155 c 208 g 260 t
ORIGIN
Query Match 63.7%; Score 811.6; DB 11; Length 932;
Best Local Similarity 97.0%; Pred. No. 5,7e-188;
Matches 859; Conservative 0; Mismatches 24; Indels 3; Gaps 3;
QY 270 CTACAAAGACCCCTTTAAACCTTATTAACAGATGTGAAAAAGAAACCTTGTG 329
Db 1 CTACAAAGACCCCTTTAAACCTTATTAACAGATGTGAAAAAGAAACCTTGTG 60
QY 330 TTGGCAATTTGTTCCGTTATTAAGATATATCTGGAACCTATGTCATCCCTGACT 389
Db 61 TTGGCAATTTGTTCCGTTATTAAGATATATCTGGAACCTATGTCATCCCTGACT 120
QY 390 GGAAGACCCAGGACAGATGATTAACATGATGCTGCTGTTGTGAGCAATGACCAATTG 449
Db 121 GGAAGACCCA -GGCAATGATTAACATGATGCTGCTGTTGTGAGCAATGACCAATTG 179
QY 450 ATGTGTGTAATTTGGAAGCAAGTATGTCGAAGGTGGAATTAATTTGGCTGAAAGTTC 509
Db 180 ATGTGTGTAATTTGGAAGCAAGTATGTCGAAGGTGGAATTAATTTGGCTGAAAGTTC 239
QY 510 TAGCATATTTGCTATGATTAAGCAAGGGGAACCGACTGGAAGTCAATTCCTTAATG 569
Db 240 TAGCATATTTGCTATGATTAAGCAAGGGGAACCGACTGGAAGTCAATTCCTTAATG 299
QY 570 TGGATGATTCCTGATGACGCCAATTAATATATGATTAATCAATGATGCAAAAGGCTGAACCTG 629
Db 300 TGGATGATTCCTGATGACGCCAATTAATATATGATTAATCAATGATGCAAAAGGCTGAACCTG 359
QY 630 GCTACTTAGAAGCTACTGCTGCTTTTGAAGAGTATTAAGTTCCTGATGGAACACG 689
Db 360 GCTACTTAGAAGCTACTGCTGCTTTTGAAGAGTATTAAGTTCCTGATGGAACACG 419
QY 690 AAATGATTTGCTGCTTAATTCAGAAATTTAAGTAAGGACTTGGCCATGATATTTA 749
Db 420 AAATGATTTGCTGCTTAATTCAGAAATTTAAGTAAGGACTTGGCCATGATATTTA 479
QY 750 AAAGCACTCATGACATTTGAAGCAATTTAGTATGATGAAGAAAGCAATGAAGCA 809
Db 480 AAAGCACTCATGACATTTGAAGCAATTTAGTATGATGAAGAAAGCAATGAAGCA 539
QY 810 GTTCATGATTAACATTTGCTGAGAGCCCTTCAAGTGTGATCTGATCTGCTGCCAG 869
Db 540 GTTCATGATTAACATTTGCTGAGAGCCCTTCAAGTGTGATCTGATCTGCTGCCAG 599
QY 870 CCATTGTGATGCTTTACACACACCTGCTGATGCTGCTGCAAGTACCAAGACGCTGG 929
Db 600 CCATTGTGATGCTTTACACACACCTGCTGATGCTGCTGCAAGTACCAAGACGCTGG 659
QY 930 ATAGGTGCTTCATCACCAGAAAGCAATGATGATTTCTGGAATACACCTGATATTG 989
Db 660 ATAGGTGCTTCATCACCAGAAAGCAATGATGATTTCTGGAATACACCTGATATTG 719
QY 990 CTACATGCTGCTTCATGAGATTTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1048
Db 720 GTACATGCTGCTTCATGAGATTTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 779
QY 1049 ATTGTGAACTCATTAACATTAAGTAAATTTCTGCTGACTATTCACATTAACGAA 1108
Db 780 ATTGTGAACTCATTAACATTAAGTAAATTTCTGCTGACTATTCACATTAACGAA 839
QY 1109 TGTATTCATCTAAGCATTTTTCATATCTCACTAAGATTAACCTTT 1154

```

Db 840 GTTATCCATCTAAGCAATTTTCATAT -TCAAGTAAGATAACTTT 884

[illegible]

	http://fulllength.invitrogen.com"			
BASE COUNT	261 a	146 c	193 g	218 t
ORIGIN	8 others			

Query Match:	60.9%;	Score 776.8;	DB 10;	Length 826;
Best Local Similarity	97.0%;	Pred. No. 1.8e-179;		
Matches 803; Conservative	8;	Mismatches 15;	Indels 2;	Gaps 2;

OY	41	TGCGGGCTGGTGGCTCTGTGGCACCAGGCGGCGGAGACACACGACATATAGACGGCTTC	100
	1	TGCGGGCTGGTGGCTCTGTGGCAGGCGGCGGAGACACGACATATAGACGGCTTC	60
Db			
OY	101	AGCACCGAGGAGCGCGCGCGCCCTTCTCCCTGGAGTACCGAGTCTTCTCCAAAATGAG	160
Db	61	AGCACCGAGGAGCGCGCGCGC - SCCTTTCCTCGAGTACCGAGCTTCTCCAAAAGAG	119
OY	161	AAGGACATATATATCTCCATTCATCATATATTCCAATTTATGAGATTAAGAGATGTTT	220
Db	120	AAGGACATATATATCTCCATTCATCATATTTCCAAATTTATGAGATTAAGAGATGTTT	179
OY	221	CACATGTAGTTGAGATACCAAGCTGTCTATATGCAAAAATGAGATTTGCTACAAAGAC	280
Db	180	CACATGTAGTTGAGATACCAAGCTGTCTATATCAAAAAAGAGATTGCTACAAAGAC	239
OY	281	CCTTTAAACCCATTAAACAAGATGTGAAAAAGAAAACTTCCTATGTTCGCAATTTG	340
Db	240	CCTTTAAACCCATTAAACAAGATGTGAAAAAGAAAACTTCCTATGTTCGCAATTTG	299
OY	341	TTTCCGGTATAAAGATATATCTGGAACATATGTGCAATCCCTGACACTTGGGAAGACCA	400
Db	300	TTTCCGGTATAAAGATATATCTGGAACATATGTGCAATCCCTGACACTTGGGAAGACCA	359

QY	401	GGCACAATGATTAACAACTACGCGCTGTGGTGAGACAATACCAATTCATGATGTGTGAA	460
Db	360	GGGCACCAATGTATAACACTACTGGCTGTGGTGAGACAATTAACCAATTCATGATGTGTGAA	419
QY	461	ATTGGAACCAAGTAGTATGTGCAAGAGGTGAATAATTTGGCTGAAAGTTCTAGCAGATTTG	520
Db	420	ATTGGAACCAAGTAGTATGTACAA-AGGTGAAAAAAATTCGCTTAAAAATTCAGGCAATATTTG	478
QY	521	GCTAATGATTGCAAGAAAGGGAAACCGACATGCGAAAGTCATTCGCCATTAATGTGATGATTCCT	580
Db	479	GCTAATAATTAACGAAGGGGAACCCGACTGGAAATCATTTGCCATTAATGTGTGATGATTCCT	538
QY	581	GATGCAAGCCAAATTAAATGATATCAATGATGTCAAACGCTGAAACCTGGCTACTTAGAA	640
Db	539	GATGCAAGCCAAATTAAATGATATCAATGATGTAAAGCGCTGAAAACTGGCTACTTAGAA	598
QY	641	GCTACTGTGCACTGGTTTAGAAGGATATPAGGTTCCGTATGAGAAAAACAGAAAAATGAGTTT	700
Db	599	GCTACTGTGCACTGGTTTAGAAGGATATPAGGTTCTCTGTGAGAAAAACAGAAAAATGAGTTT	658
QY	701	GGCTTAATGACGATTTAAAGATPAAGCACTTTCGCATGTATTTATTAAGACATCAT	760
Db	659	GGCTTAATGACGATTTAAAGATPAAGCACTTTCGCATTTATTTATTAAGACATCAT	718
QY	761	GACCAATTGGAAGCACTTGTGACTPAGAAAAACGAATGGAAGAAAGCAATCAGTGCATGAT	820
Db	719	GACCAATTGGAAGCACTTGTGATTTAGTAAACGATGGAAGAAAGCAATTAAGTGCATGAT	778
QY	821	ACAACCTTGTCTGAGAGCCCTTCACAAGTATATCCATGCTGTGCCAGA	868
Db	779	ACAACCTTGTCTGAGAGCCCTTCACAAGTATATCCATGCTGTGCCAGA	826

RESULT		7			
BG674343					
Locus					
DEFINITION					
.ACCESSION	BG674343	996 bp	mRNA		01-MAY-2001
VERSION	60262032SF1	NCL_CGNF_Skn3	Homo sapiens cDNA clone IMAGE:4745743	5	
KEYWORDS	BG674343		mRNA sequence.		
SOURCE	BG674343.1	GI:13905739			
ORGANISM	EST .				
	human.				
	Homo sapiens				
	Euryarchaea; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Euarcharia; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 996)				
AUTHORS	Nih-MGC http://mgc.ncl.nih.gov/.				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished. (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D.				

BASE COUNT
ORIGIN

Query Match 59.6%; Score 760; DB 11; Length 996;
 Best Local Similarity 89.6%; Pred. No. 2.4e-175;
 Matches 896; Conservative 0; Mismatches 95; Indels 9; Gaps 7;

```

OY 39 CCGCGGGGCTGCTGCTCTGTGACGCGGCGGCGGAGAGACTCCGACATATGAGCGGT 98
    1 CCGCGGGGCTGCTGCTCTGTGACGCGGCGGCGGAGAGACTCCGACATATGAGCGGT 60
OY 99 TCAGCAGCAGGAGGCGCGCGCGCTTCTCCTGAGATGACGAGTCTTCTCAAAATG 158
    61 TCAGCAGCAGGAGGCGCGCGCGCTTCTCCTGAGATGACGAGTCTTCTCAAAATG 120
OY 159 AGAAGAGACAAATATATCTCCATTCATGATATTCATTTTGAAGATGAGTGT 218
    121 AGAAGAGACAAATATATCTCCATTCATGATATTCATTTTGAAGATGAGTGT 180
OY 219 TTCACATGATAGTGAAGTACGACGCTGCTTAATGCAAAATGAGATGCTACAAAG 278
    181 TTCACATGATAGTGAAGTACGACGCTGCTTAATGCAAAATGAGATGCTACAAAG 240
OY 279 ACCCTTTAAACCTATTTAAACAAGATGTGAAAAAGAAAACCTGCTATGTCGAAT 338
    241 ACCCTTTAAACCTATTTAAACAAGATGTGAAAAAGAAAACCTGCTATGTCGAAT 300
OY 339 TGTTCCTGATTAAGATATATCTGCAACTATGTCGATCCTCAGACTTGGGAAGAC 398
    301 TGTTCCTGATTAAGATATATCTGCAACTATGTCGATCCTCAGACTTGGGAAGAC 360
OY 399 CAGGGGACAATGATTAACATATCTGCTGTGTGTCGATGACCAATGATGTGTG 458
    361 CAGGGGACAATGATTAACATATCTGCTGTGTGTCGATGACCAATGATGTGTG 420
OY 459 AATTTGGAAGCAAGTATGTGCAAGAGTGAATTAATGCGTGAAGTCTGAGCAT 518
    421 AATTTGGAAGCAAGTATGTGCAAGAGTGAATTAATGCGTGAAGTCTGAGCAT 480
OY 519 TGCTATGATTTGAGAAAGGGAACGACATGGAAGATCATTCGATTAATGATGATG 578
    481 TGCTATGATTTGAGAAAGGGAACGACATGGAAGATCATTCGATTAATGATGATG 540
OY 579 CTGATGACGCAATTTATATGATATCATGATGTCAAACGCTGAAACCTGCTACTT 638
    541 CTGATGACGCAATTTATATGATATCATGATGTCAAACGCTGAAACCTGCTACTT 600
OY 639 AACCTACTGTGACT-GGTTTGAAGATATAGCTTCTGATGGAAGAACCAATGAG 697
    601 AACCTACTGTGACTGGGTTTGAAGATATAGCTTCTGATGGAAGAACCAATGAG 660
OY 698 TTTGCGTTTATGACAGATTTAAAGATTAAGGACTTTCATGATATTTAAAGCAT 757
    661 -TTGCGTTTATGACAGATTTAAAGATTAAGGACTTTCATGATATTTAAAGCAT 719
OY 758 CATGACCAATTTGAAGCAATTAAGTACTAAGAAAACGAATGGA-AAAGAAATCA 816
    720 TCATGACCAATTTGAAGCAATTAAGTACTAAGAAAACGAATGGAATCAATG 779
OY 817 GAATTAACACTTGTGTGAGAGCCCTTCAAGTGTGATCCTGATGCTGCGAGGACAT 876
    780 GAATTAACACTTGTGTGAGAGCCCTTCAAGTGTGATCCTGATGCTGCGAGG 835
OY 877 GATGCTTTTACCAACCCGTGATGCTGACAGTACCAACCAAGATGATGATAGTG 936
    836 GATGCTTTTACCAACCCGTGATGCTGACAGTACCAACCAAGATGATGATAGTG 893
OY 937 GTTTCATCACCAGAAAACATTAATGATTTCTGTAATCAAGCTGATTTGCTACAT 996
    894 GTTTCATCACCAGAAAACATTAATGATTTCTGTAATCAAGCTGATTTGCTACAT 953
OY 997 GTGTTTCTGATGATTAATTAAGATTAAGTAAAGTAAAGTAAAGTAAAGTAA 1036
    954 GGAGTTTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 993
  
```

RESULT 8
 BG715195
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

8
 BG715195 773 bp mRNA
 602675655F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4798285 5',
 mRNA sequence.
 BG715195
 BG715195.1 GI:13994274
 EST.
 human
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 773)
 NIH-MGC <http://mgi.ncl.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLM10685 Row: n Column: 14
 High quality sequence stop: 773.
 Location/Qualifiers
 1..773
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4798285"
 /clone_lib="NIH_MGC_96"
 /tissue_type="hypothalamus"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescript (modified
 pBluescript KS+); Site_1: BamHI; Site_2: SalI; XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTAA-3',
 size-selected for average insert size 2.3 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 251 a 137 c 184 g 201 t
 ORIGIN

Query Match 58.3%; Score 742.8; DB 11; Length 773;
 Best Local Similarity 99.5%; Pred. No. 3.6e-171;
 Matches 766; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

```

OY 84 GCACTATGACGCGCTTACAGACGAGGAGCGCGCCCTTCTCCCTGAGTACCGAG 143
    4 GCACTATGACGCGCTTACAGACGAGGAGCGCGCCCTTCTCCCTGAGTACCGAG 63
OY 144 TCTTCTCAAAAATGAGAAAGACATATATCTCCATTTGATGATTTCAATTTATG 203
    64 TCTTCTCAAAAATGAGAAAGACATATATCTCCATTTGATGATTTCAATTTATG 123
OY 204 CAGATTAAGATGCTTACATGATGATGATGATGATGATGATGATGATGATGATG 263
    124 CAGATTAAGATGCTTACATGATGATGATGATGATGATGATGATGATGATGATG 183
OY 264 AGATTGCTACAAAGACCCCTTAACCCCTTAATTAACAAGATGTGAAAAAGAAATTC 323
    184 AGATTGCTACAAAGACCCCTTAACCCCTTAATTAACAAGATGTGAAAAAGAAATTC 243
OY 324 GCTATGTTGGAAATTTGCTCCGATTAAGGATATATCTGGAATGATGATGATGAT 383
    GCTATGTTGGAAATTTGCTCCGATTAAGGATATATCTGGAATGATGATGATGATGAT 383
  
```

Db 244 GGTATGTTGGCAATTTGTTCCGGTAAAGATATATGGAATATGAGTATGTCATCCCTC 303
OY 384 AGACTTGGAGAGCCAGGAGGACAAATGATTAACATCTGCTGTTGTGTGACATGACC 443
Db 304 AGACTTGGAGAGCCAGGAGGACAAATGATTAACATCTGCTGTTGTGTGACATGACC 363
OY 444 CAATGATGATGTGTAATTTGGAAAGGATGTCGAAGAGCTGAATTAATTTGGCGTGA 503
Db 364 CAATGATGATGTGTAATTTGGAAAGGATGTCGAAGAGCTGAATTAATTTGGCGTGA 423
OY 504 AAGTCTAGGAGCAATTTGGCTATGATTTGACGAAGGGGAAACCGATGGAAGTCAATGCCA 563
Db 424 AAGTCTAGGAGCAATTTGGCTATGATTTGACGAAGGGGAAACCGATGGAAGTCAATGCCA 483
OY 564 TTAATGATGATGATGCTGATGAGCAATTTAATGATATCAATGATGCAAGGGCTGA 623
Db 484 TTAATGATGATGATGCTGATGAGCAATTTAATGATATCAATGATGCAAGGGCTGA 543
OY 624 AACCTGGCTACTTAGAGCTGCTGAGCTGTTAGAAAGTATTAAGTTCCTGATGGA 683
Db 544 AACCTGGCTACTTAGAGCTGCTGAGCTGTTAGAAAGTATTAAGTTCCTGATGGA 603
OY 684 AACCGAAATAGATTTGCTGTTAATGAGATTT-AAAGATAGGACTTTGCCATTTGAT 742
Db 604 AACCGAAATAGATTTGCTGTTAATGAGATTTAAAGATGAGACTTTGCCATTTGAT 663
OY 743 ATTATTAAGACATGACATGACATGGAAGCATTTAGTACTAGAGAAACGAATGAA-A 801
Db 664 ATTATTAAGACATGACATGACATGGAAGCATTTAGTACTAGAGAAACGAATGAA-A 723
OY 802 AGCAATCAAGTTCATGATATACATCACTTTGCTGAGAGCCCTTCAGAGTGTG-851
Db 724 AGCAATCAAGTTCATGATATACATCACTTTGCTGAGAGCCCTTCAGAGTGTG-773

RESULT 9
B1258591 771 bp mRNA EST 17-JUL-2001
LOCUS 602969219P1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5109100 5'
DEFINITION mRNA sequence.
B1258591
ACCESSION B1258591.1 GI:14815097
VERSION EST.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 771)
NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
AUTHORS Email: cgaab-r@mail.nih.gov
CONTACT: Robert Strusberg, Ph.D.
COMMENT Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM11264 row: e column: 05
High quality sequence stop: 749.
Location/Qualifiers
1..771
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5109100"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.4 kb. Library prepared by Life

BASE COUNT 264 a 132 c 155 g 220 t
ORIGIN
Technology: "Best Local Similarity 99.3%; Pred. No. 5,7e-171; Matches 745; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Query Match 58.2%; Score 742; DB 11; Length 771;
Best Local Similarity 99.3%; Pred. No. 5,7e-171; Matches 745; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 521 GCTATGATTTGACGAAGGGAACCGACTGGAAGCATGCTGCAATTAATGATGATGATCCT 580
Db 1 GCTATGATTTGACGAAGGGAACCGACTGGAAGCATGCTGCAATTAATGATGATGATCCT 60
OY 581 GATGAGCCCAATTTAAAGATATCAATGATGTCGAACCGCTGGAACCTGCTACTAGAA 640
Db 61 GATGAGCCCAATTTAAAGATATCAATGATGTCGAACCGCTGGAACCTGCTACTAGAA 120
OY 641 GCTACTGAGCTGCTTTAGAAAGTATTAAGTTCCTGATGGAAGAAACCGAAATGAGTTT 700
Db 121 GCTACTGAGCTGCTTTAGAAAGTATTAAGTTCCTGATGGAAGAAACCGAAATGAGTTT 180
OY 701 GCGTTAATGAGCAATTTAAAGATTAAGAGCTTTGCCATTTGATTAATTAAGACATCAT 760
Db 181 GCGTTAATGAGCAATTTAAAGATTAAGAGCTTTGCCATTTGATTAATTAAGACATCAT 240
OY 761 GACCAATGGAAGCAATTTAGTACTAGAAACGAATGGAAGAAAGCAATGTCATGAT 820
Db 241 GACCAATGGAAGCAATTTAGTACTAGAAACGAATGGAAGAAAGCAATGTCATGAT 300
OY 821 ACAATTTGCTGAGAGCCCTTCAGAGTGTGATGCTGCTGCAAGGCAATGCTGAT 880
Db 301 ACAATTTGCTGAGAGCCCTTCAGAGTGTGATGCTGCTGCAAGGCAATGCTGAT 360
OY 881 GCTTTACCAACCACTGTAAGTCTGCTGCAAGTACAGAGAGCTGATAGAGTTC 940
Db 361 GCTTTACCAACCACTGTAAGTCTGCTGCAAGTACAGAGAGCTGATAGAGTTC 420
OY 941 CATCACCAAGAAACTAATGAGATTTCTGTAATACAGCTGATTTGCTACATGCT 1000
Db 421 CATCACCAAGAAACTAATGAGATTTCTGTAATACAGAGCTGATTTGCTACATGCT 480
OY 1001 TCATCTGATGATTTAGAAAGTAAAGTAAAGTAAAGCTTTTAAAGCTTTAAATTTGTAAGT 1060
Db 481 TCATCTGATGATTTAGAAAGTAAAGTAAAGTAAAGCTTTTAAAGCTTTAAATTTGTAAGT 540
OY 1061 CATCTAAGTAAGTAATTTCTGCTGCTGCTAATCAATCAATCAATCAATCAATCAAT 1120
Db 541 CATCTAAGTAAGTAATTTCTGCTGCTGCTAATCAATCAATCAATCAATCAATCAAT 600
OY 1121 AAAGATTTTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 1180
Db 601 AAAGATTTTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 660
OY 1181 TTGCTATTTGGAAGTCACTTTGTAATAGATGTCGAAGGGGACACATATTTGATATAT 1240
Db 661 TTGCTATTTGGAAGTCACTTTGTAATAGATGTCGAAGGGGACACATATTTGATATAT 720
OY 1241 GTTACATATGTTAGAAATTAATTTT 1270
Db 721 GTTACATATGTTAGAAATTAATTTT 750

RESULT 10
BG713976 772 bp mRNA EST 08-MAY-2001
LOCUS 602674374P1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4796875 5'
DEFINITION mRNA sequence.
BG713976
ACCESSION BG713976.1 GI:13992907
VERSION EST.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 772)	Mammalia: Eutheria: Primates; Catarrhini; Homnidae: Homo.	NIH-MGC http://mgc.nci.nih.gov/	National Institutes of Health, Mammalian Gene Collection (MGC)	unpublished (1999)
Contact:	Robert Strausberg, Ph.D.			

BASE COUNT	235 a	146 c	197 g	194 t
ORIGIN				

Query Match	58.0%	Score 739.8	DB 11	Length 772
Best Local Similarity	99.5%	Pred. No. 2e-170		
Matches 763	Conservative	0	Mismatches 2	Indels 2
				Gaps 2

QY	28	AGTGGGCGCGGCGTGGCGGCGTGGCGTGGCGAGCGCGCGGCGAGACGTCGGGAC	87
Db	6	AGTGGCGCGCGCGTGGCGGCGTGGCGTGGCGAGCGCGCGGCGAGACGTCGGGAC	65
QY	88	TATGAGCGGCTTCAGCACCGAGAGCGCGCGCGCGCTTCCTCGTGGATACGACTCTT	147
Db	66	TATGAGCGGCTTCAGCACCGAGAGCGCGCGCGCGCTTCCTCGTGGATACGACTCTT	125
QY	148	CCTCAAAAATATGAAAGAGACATATATATCTTCATTTCTCATTTATTCGAATTTATGACA	207
Db	126	CCTCAAAAATATGAAAGAGACATATATATCTTCATTTCTCATTTATTCGAATTTATGACA	185
QY	208	TAAAGATGTGTTTCACATGTAGTGAAGTACACGCGTGGTAAATGCAAAATGGAGAT	267
Db	186	TAAAGATGTGTTTCACATGTAGTGAAGTACACGCGTGGTAAATGCAAAATGGAGAT	245
QY	268	TGCTACAAAGACCCCTTTAAACCCCTATTAAACAGATGTGAAAAAGAAAACTTCGCTA	327
Db	246	TGCTACAAAGACCCCTTTAAACCCCTATTAAACAGATGTGAAAAAGAAAACTTCGCTA	305
QY	328	TGTTGCGAATTTGTTCCCGTATAAAGGATATATCTGGAACATATGGTGGCATCCCTCAGAC	387
Db	306	TGTTGCGAATTTGTTCCCGTATAAAGGATATATCTGGAACATATGGTGGCATCCCTCAGAC	365
QY	388	TTTGGGAAGACCCAGGGACCAATGATAAACAATCTGGCTTGTGGTGACAATGACCCAAAT	447
Db	366	TTTGGGAAGACCCAGGGACCAATGATAAACAATCTGGCTTGTGGTGACAATGACCCAAAT	425
QY	448	TGATGTGTGTGAATTTGGAAGCAGGTATGTGCAAGAGGTGCAAAATATTTGGCGTGAAGT	507
Db	426	TGATGTGTGTGAATTTGGAAGCAGGTATGTGCAAGAGGTGCAAAATATTTGGCGTGAAGT	485

OY	508	TC TAGGCATATTTGGCTATGATTTAGACGAAGGGGAAACCGCATGTGAAGTCATTTGCCATTAA	567
Db	486	TC TAGGCATATTTGGCTATGATTTAGACGAAGGGGAAACCGCATGTGAAGTCATTTGCCATTAA	545
OY	568	TGTGATGATCCTCGATGTCAGCCCAATTATATATCATGTATGTCAAAAGGCTGAAAC	627
Db	546	TGTGATGATATCTCGATGTCAGCCCAATTATATATCATGTATGTCAAAAGGCTGAAAC	605
OY	628	TGGCCTACTTAGACGATCTGTGACTGGTTTGTAAGATATTAAGTTCCTGATGAAAAAC	687
Db	606	TGGCCTACTTAGACGATCTGTGACTGGTTTGTAAGATATTAAGTTCCTGATGAAAAAC	665
OY	688	AGAAATATG - AGTTTGCCTTTAATGCGAATTTAAAGATPAGACATTTGCCATTGA - TATT	745
Db	666	AGAAATATGAGTTTGCCTTTAATGCGAATTTAAAGATPAGACATTTGCCATTGACTATT	725
OY	746	ATTAAAGCAGCATGATGACCATTTGGAAGACATTAAGCATTAAGAAAC	792
Db	726	ATTAAAGCAGCATGATGACCATTTGGAAGACATTAAGCATTAAGAAAC	772

RESULT	11
LOCUS	B1259308
DEFINITION	B1259308 878 bp mRNA EST 17-JUL-2001 60297254.6F1 NIH_MGC_12 Homo sapiens CDNA clone IMAGE:5111963 5'
ACCESSION	B1259308
VERSION	B1259308
KEYWORDS	B1259308.1 GI:14816515
SOURCE	EST.
ORGANISM	human;
	Homo sapiens
	Euchariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 878)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

High quality sequence stop: 78%.

BASE COUNT	295 a	141 c	192 g	250 f
ORIGIN				

Query Match	57.5%	Score 732.8	DB 11	Length 878
Best Local Similarity	96.1%	Pred No 1e-158		
Matches 828	Conservative	0	Mismatches 22	Indels 12
				Gaps
OY	423	GCTGTGTGGTGACATGACCCCATTTGATGTGTGTGGAATTTGGAGCAAGTATGTGC	AA	482
db	1	GCTGTGTGGTGAC-ATACCCCAATTTGATGTGTGGAATTTGGAGCAAGTATGTGC	AA	59


```

/db_xref="MGI:1915145"
/clone="9130233G20"
/sex="male"
/issue_type="cecum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
112..591
CDS
/feature="putative"
/codon_start=1
/protein_id="BAB32057.1"
/db_xref="GI:12860833"
/translation="MSGF5SERAPFLEVRFLKNEKGOYISPHVDPIYADKDF
HMVYVPRMSNAKMEIATKDPNPKOVKKKLYVANLPYKRYINNYGAIPTWE
DPGSHDKAGCGDNDPIDVCEIGSKVARSITIKVKVLGIAMIDEGEDRLEGCH"
1291..1296
polyA_signal
/feature="putative"
1308
polyA_site
/feature="putative"
293 c 338 g 319 t
BASE COUNT 358 a 293 c 338 g 319 t
ORIGIN
Query Match 57.3% Score 730.2; DB 12; Length 1308;
Best Local Similarity 80.4%; Pred. No. 5e-168;
Matches 1009; Conservative 0; Mismatches 218; Indels 28; Gaps 12;
QY 40 GTGCGGGCTGTGCTGTGTCGAGCGGCGGAGACTCGGCACTATGAGCGGCTT 99
DB 63 GTGTGTTGTGTGCTGTGTCGAGCGGCGGAGACTCGGCACTATGAGCGGCTT 122
QY 100 CAGCACCGAGAGCGCGCGCGCTTCCCTGAGTACCGAGCTTCTCAAAAATGA 159
DB 123 CAGCACCGAGAGCGCGCGCGCTTCCCTGAGTACCGAGCTTCTCAAAAATGA 182
QY 160 GAAAGGACATATATATTCATTCATATTCATATTCATATTCATATTCATATTCAT 219
DB 183 GAAAGGACATATATTCATTCATATTCATATTCATATTCATATTCATATTCATATTCAT 242
QY 220 TCACATGTAGTGTGAAGTACCAAGCTGTCTATGCAAAAATGAGATGTGTACAAGA 279
DB 243 CCACATGTGTGTGAGTGTGTGTCGAGCGGCGGAGACTCGGCACTATGAGCGGCTT 302
QY 280 CCCTTTAAACCTATTTAAACAAGATGTGAAGAAAAGAAACCTTGGTATGTGCGAATTT 339
DB 303 CCCTTTAAACCTATTTAAACAAGATGTGAAGAAAAGAAACCTTGGTATGTGCGAATTT 362
QY 340 GTTCCCGTAAAGATATATTCATTCATATTCATATTCATATTCATATTCATATTCAT 399
DB 363 GTTCCCGTAAAGATATATTCATTCATATTCATATTCATATTCATATTCATATTCATATTCAT 422
QY 400 AGGCGACATGTAAACACTACTGCTGTGTGTGACAAATGACCCCAATGATGTGTGTGA 459
DB 423 AGGCGACATGTAAACACTACTGCTGTGTGTGACAAATGACCCCAATGATGTGTGTGA 482
QY 460 AATGTGAACAAGTATGTGCAAGAGTGAATATTTGGCGTGAAGTTCTAGCAATATT 519
DB 483 AATGTGAACAAGTATGTGCAAGAGTGAATATTTGGCGTGAAGTTCTAGCAATATT 542
QY 520 GCGTATGATGTGCAAGAGGAA-ACCGAGTGAAGTCAATGCAATTAATGTG-ATGA 576
DB 543 GCGTATGATGTGCAAGAGGAA-ACCGAGTGAAGTCAATGCAATTAATGTG-ATGA 602
QY 577 TCGTGTATG-AGCCAAATTAATATGATA-TCAATGATGTGCAAGGCTGA-CTG 630
DB 603 TCGTGTATG-AGCCAAATTAATATGATA-TCAATGATGTGCAAGGCTGA-CTG 662
QY 631 CTACTTAAAGTACT-CTGAGACTGTGTAGAAAGTATAG-TCCTGATGAAAGC 687
DB 663 CTACTTAAAGTACT-CTGAGACTGTGTAGAAAGTATAG-TCCTGATGAAAGC 722
QY 688 AAAAAATGTTGCTTAAATGCAAGATTAAGATAGGACTTGCATGATATTTAT 747
DB 723 AAAAAATGTTGCTTAAATGCAAGATTAAGATAGGACTTGCATGATATTTAT 782

```

```

QY 748 TAAAGCACTCATGACCATTGGAAGCACTTAGTACTAGAAAAAGCAAGTAAAGCAAT 807
DB 783 TAAAGCACTCATGACCATTGGAAGCACTTAGTACTAGAAAAAGCAAGTAAAGCAAT 842
QY 808 CAGTTCATGATATACAACTTTGTCTGAGAGCCCTTCAAGTGTGATCTGATCTGCGAG 867
DB 843 CAGTTCATGATATACAACTTTGTCTGAGAGCCCTTCAAGTGTGATCTGATCTGCGAG 902
QY 868 AGCCATTGTGATGCTTTTACACACACCCCTGTGATGCTGCTGACAGTACCAAGCAAGCT 927
DB 903 AGCCATTGTGATGCTTTTACACACACCCCTGTGATGCTGCTGACAGTACCAAGCAAGCT 962
QY 928 GATTAAGTGTTCATACACCAAGAACTAATGAGATTTCTGTGATATACAGCTGATAT 987
DB 963 GATTAAGTGTTCATACACCAAGAACTAATGAGATTTCTGTGATATACAGCTGATAT 1022
QY 988 TGTACATGCTGTTCATCTGATGATGTATGAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1047
DB 1023 GCGCGCATCTCTGTTCACCTGGAAGCATTTAGTAAAGCAAGCAAGCAAGCAAGCAAG 1076
QY 1048 AATTTGTAGACATCATCTACTACTAAAGTAAATTCGCTGATCAATCAATATATGACAG 1107
DB 1077 AATTTGTAGACATCATCTACTACTAAAGTAAATTCGCTGATCAATCAATATATGACAG 1135
QY 1108 ATGTTATCCATCTAAAGCAATTTT-----TCATATCTCAATAGATATCTTTAGCAC 1160
DB 1136 CTGCTTCAGAGGTATGACCATCTAGACACCTCATCTCAATCAAGCAAGCAAGCTTTATTA 1194
QY 1161 ATGCTTAATATCAAGCAAGTGTCTATTTGGAAGTCACTTGTGAATATGATGTCAAGGG 1220
DB 1195 AACCTTAGCTCTCAAGCGGTGGGACATGGAAGTATTTGTCATGATGTGCTGCTGCTG 1254
QY 1221 AGCACAATTTGATGTATATGTTATGATATGTTAGGAATTAATATTTGCTG 1275
DB 1255 GACACCTATTTGATGTATGATGTTATGATGTTATGTTAGGAATTAATATTTGCTG 1308

```

```

RESULT 13
LOCUS BG574468 915 bp mRNA EST 10-APR-2001
DEFINITION 602596421F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4705253 5',
ACCESSION BG574468
VERSION BG574468.1 GI:13582121
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DPF
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHM10571 row: 3 column: 06
High quality sequence stop: 830.
Location/Qualifiers
1..915
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4705253"
/clone_lib="NIH_MGC_87"
/issue_type="mammalian adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"

```

```

FEATURES
source

```



```

FEATURES
  source
    location/Qualifiers
      1. .762
        /organism="Homo sapiens"

```

OY	23	GTCAATCGGGCCCGGTCGCGGGCTGGTGGCTCTGGAGGGGCGGAGGAGCTCGG	85
Db	6	GTCAATCGGGCCCGGTCGCGGGCTGGTGGCTCTGGAGGGGCGGAGGAGCTCGG	65
OY	85	CACATAGGCGGGCTTCACACCGAGGAGCGCGCGGCCCTTCCTCCCTGGAGTACCGAGT	144
OY	145	CTTCTCTCAAAATGAGAAAGGACATATATATCTCCATTTTCATGATATTTCCATTTATGC	204
Db	126	CTTCTCTCAAAATGAGAAAGGACATATATATCTCCATTTTCATGATATTTCCATTTATGC	185
OY	205	AGATATAGATGTGTTTCACATGTAAGTGAAGTACACGCTGGTCTAATGCAAAATGGA	264
Db	186	AGATATAGATGTGTTTCACATGTAAGTGAAGTACACGCTGGTCTAATGCAAAATGGA	245
OY	265	GATTCTCTCAAAAGGACCCCTTTAAACCCTATTTAAACAAGATGTGAAAAAGAAACCTTCG	324
Db	246	GATTCTCTCAAAAGGACCCCTTTAAACCCTATTTAAACAAGATGTGAAAAAGAAACCTTCG	305
OY	325	CTATGTTCCGAATTTGTTGTTCCCGTATTAAGGATATATCTGGAACATGTGGTCCATCCCTCA	384
Db	306	CTATGTTCCGAATTTGTTGTTCCCGTATTAAGGATATATCTGGAACATGTGGTCCATCCCTCA	365
OY	385	GACTTGGGAGAGCCAGGCGCAATGATAAACAATCTGGCTGTTGTGTGACAAATGACCC	444
Db	366	GACTTGGGAGAGCCAGGCGCAATGATAAACAATCTGGCTGTTGTGTGACAAATGACCC	425
OY	445	AATGATGTGTGTGAATTTGGAACCAAGGTATGTGCAAGAGGTGAATTAATTTGGCCGGA	504
Db	426	AATGATGTGTGTGAATTTGGAACCAAGGTATGTGCAAGAGGTGAATTAATTTGGCCGGA	485
OY	505	AGTTCTAGCATATTTGGCTATGATGATTGACGAAGGGGAAACGACATGGAAGTCATTGCCAT	564
Db	486	AGTTCTAGCATATTTGGCTATGATGATTGACGAAGGGGAAACGACATGGAAGTCATTGCCAT	545
OY	565	TAAATGTGATGATCTCTGATGCAGCCAATTATATGATATCAATGATGTCAAAGGCTGGA	624
Db	546	TAAATGTGATGATCTCTGATGCAGCCAATTATATGATATCAATGATGTCAAAGCG -TGAA	604
OY	625	ACCTGCTACTTAGAAGGTACTGTGGAGCTGTTTGAAGGATATTAAGTTCCTGATGGA	684
Db	605	ACCTGCTACTTAGAAGGTACTGTGGAGCTGTTTGAAGGATATTAAGTTCCTGATGGA	664
OY	685	ACCAGAAATAGATTGGCTTTAATGCAGAATTTAAAGATTAAGGACTTGGCATTTGATAT	744
Db	665	ACCAGAAATAGATTGGCTTTAATGCAGAATTTAAAGATTAAGGACTTGG -CATTTGAT	722
OY	745	TATTTAAAGCACTATGACCAATTTGGAAGCAATTAGTGA 782	
Db	723	TATTTAAAGCACTATGACCAATTTGGAAGCAATTAGTGA 760	

Mon Apr 1 07:59:58 2002

us-09-415-540-2.std.rst

Search completed: March 29, 2002, 19:50:17
Job time: 3420 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2002, 19:50:25 ; Search time 149.95 seconds

(without alignments)
9134.216 Million cell updates/sec

Title: US-09-415-540-2
Perfect score: 1275
Sequence: 1 CAAGAGGTTGGCGCTCTCT.....GAATTAATAATTATTTGCTG 1275

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST: *
1: em_estfun: *
2: em_esthum: *
3: em_estlin: *
4: em_estlpl: *
5: em_estpl: *
6: em_estba: *
7: em_estro: *
8: em_estov: *
9: em_hic: *
10: qb_estl: *
11: qb_est2: *
12: qb_hic: *
13: qb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_rtd: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	751	58.9	826	10 AL519324	AL519324 AL519324
2	709	55.6	714	2 BG700549	BG700549 602680920
3	656	51.5	771	11 B1258591	B1258591 602968519
4	636	49.9	773	11 BG715195	BG715195 602675655
5	635	49.8	905	11 BG027821	BG027821 602294777
6	632	49.6	915	11 BG574468	BG574468 602596421
7	628	49.3	628	11 BE887465	BE887465 601508111
8	625	49.0	660	10 AM009649	AM009649 w885-03.x
9	618	48.5	772	11 BG713976	BG713976 602674374
10	607	47.6	955	11 BF791373	BF791373 602251332
11	599	47.0	602	11 BG504546	BG504546 602552528
12	595	46.7	596	11 BG574238	BG574238 602596140

C	13	592	46.4	998	11 BF966851	BF966851 602286503
C	14	586	46.0	932	11 BG676397	BG676397 602622839
C	15	584	45.8	584	10 A1719667	A1719667 at4.f12.x
C	16	581	45.6	996	11 BG674343	BG674343 602620305
C	17	579	45.4	882	11 BF966192	BF966192 602286503
C	18	577	45.3	711	11 BG777143	BG777143 602664359
C	19	575	45.1	575	10 AM262869	AM262869 x996e02.x
C	20	572	44.9	714	11 BG702955	BG702955 602684915
C	21	570	44.7	769	11 BG702317	BG702317 602683563
C	22	567	44.5	567	11 BG180287	BG180287 602331122
C	23	567	44.5	702	10 AM840924	AM840924 RC1-CN000
C	24	562	44.1	1076	11 BG179769	BG179769 602328856
C	25	557	43.7	585	10 AM873109	AM873109 hq22h11.x
C	26	555	43.5	589	10 AM469221	AM469221 hc79e09.x
C	27	552	43.3	592	11 BE867111	BE867111 601442951
C	28	544	42.7	762	11 BG721093	BG721093 602692932
C	29	543	42.6	639	10 AA694204	AA694204 z12c12.s
C	30	541	42.4	564	10 AA493425	AA493425 ne48h03.s
C	31	539	42.3	750	11 BG704401	BG704401 602687468
C	32	535	42.0	593	11 BG714352	BG714352 602669861
C	33	532	41.7	914	11 BF033412	BF033412 601457932
C	34	528	41.4	631	10 AM518147	AM518147 xx81f10.x
C	35	523	41.0	570	10 A1215898	A1215898 qm35h09.x
C	36	517	40.5	690	10 AV706813	AV706813 AV706813
C	37	516	40.5	775	11 BG036446	BG036446 602326583
C	38	515	40.4	794	10 BE379621	BE379621 601159361
C	39	510	40.0	617	10 A1818185	A1818185 w42c03.x
C	40	504	39.5	940	11 BG254456	BG254456 602369062
C	41	503	39.5	503	11 BE887538	BE887538 601508191
C	42	503	39.5	585	10 BE042554	BE042554 hc25e08.x
C	43	503	39.5	624	11 BG777846	BG777846 602664958
C	44	503	39.5	804	11 B1260339	B1260339 602969372
C	45	502	39.4	624	10 A1922084	A1922084 w88b03.x

ALIGNMENTS

RESULT 1
AL519324 826 bp mRNA EST 13-FEB-2001
LOCUS AL519324 LTI_NFL011.NBC1 Homo sapiens cDNA clone CSODAO12YD02 5
DEFINITION AL519324 LTI_NFL011.NBC1 Homo sapiens cDNA clone CSODAO12YD02 5
prime, mRNA sequence.
ACCESSION AL519324 GI:12782817
VERSION AL519324.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 826)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source
1. 826
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODAO12YD02"
/clone_lib="LTI_NFL011.NBC1"
/sex="male"
/tissue="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library is not normalized, but is the control for the normalized libraries. Library was constructed by Life

JOURNAL COMMENT

Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
 Plate: L1AM10685 row: n column: 14
 High quality sequence stop: 773.
 Location/Qualifiers

FEATURES

source

1..773
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4798285"
 /clone_lib="NIH_MGC_96"
 /tissue_type="hypothalamus"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: phluescriptR (modified
 Bluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcgaag
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3',
 size selected for average insert size 2.3 kb and
 normalized to 10⁷ 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 251 a 137 c 184 g 201 t
 ORIGIN

Query Match 49.9%; Score 636; DB 11; Length 773;
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

86 ACTATGAGCGGCTTACGACCGGAGCGCGCCCTTCTCCCTGAGTACCGAGTC 145
 Db 6 ACTATGAGCGGCTTACGACCGGAGCGCGCCCTTCTCCCTGAGTACCGAGTC 65
 QY 146 TTCCCAAAAATGAAAGGACCAATATATATCTCCATTTTCATGATTTCCATTTATGCA 205
 Db 66 TTCCCAAAAATGAAAGGACCAATATATATCTCCATTTTCATGATTTCCATTTATGCA 125
 QY 206 GATAAGATGTTTTCACATGTTAGTACCAAGCGTGTCTAATGCAAAATGAG 265
 Db 126 GATAAGATGTTTTCACATGTTAGTACCAAGCGTGTCTAATGCAAAATGAG 185
 QY 266 ATTCGTACAAAGAGCCCTTTAAACCCCTTAACAAGATGTGAAGAAAGAAACTTCGC 325
 Db 186 ATTCGTACAAAGAGCCCTTTAAACCCCTTAACAAGATGTGAAGAAAGAAACTTCGC 245
 QY 326 TATGTTGGCAATTTTTCCTCCGATATATGAGATATCTGAACTATGAGCCATCCCTAG 385
 Db 246 TATGTTGGCAATTTTTCCTCCGATATATGAGATATCTGAACTATGAGCCATCCCTAG 305
 QY 386 ACTTGGAGAGCCAGGCGACAAATGATAACATACTGGCTGTGTGTGACAAATGACCA 445
 Db 306 ACTTGGAGAGCCAGGCGACAAATGATAACATACTGGCTGTGTGTGACAAATGACCA 365
 QY 446 ATTGATGTTGTGAATTTGGAAGCAAGTATGTGCAAGAGTGAATTAATTTGGCTGAAA 505
 Db 366 ATTGATGTTGTGAATTTGGAAGCAAGTATGTGCAAGAGTGAATTAATTTGGCTGAAA 425
 QY 506 GTTCAGCATATTTGGCTATGATGACGAAGGGGAAACCGATGGAAGTCTTCCCAT 565
 Db 426 GTTCAGCATATTTGGCTATGATGACGAAGGGGAAACCGATGGAAGTCTTCCCAT 485
 QY 566 AATGTGATGATCTGATGACGACCAATTAATATGATATCAATGATGCAAAAGGCTGAAA 625
 Db 486 AATGTGATGATCTGATGACGACCAATTAATATGATATCAATGATGCAAAAGGCTGAAA 545

QY 626 CTGCTACTTAGAAGCTACTGTGACCTGTTTAAGAGTATTAAGCTTCCGATGAGAA 685
 Db 546 CTGCTACTTAGAAGCTACTGTGACCTGTTTAAGAGTATTAAGCTTCCGATGAGAA 605
 QY 686 CCAGAAATGAGTTGCGTTTATGACAGATTTAA 721
 Db 606 CCAGAAATGAGTTGCGTTTATGACAGATTTAA 641

RESULT 5

LOCUS B6027821 905 bp mRNA EST 24-JAN-2001
 DEFINITION 602294777F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4389594 5',
 mRNA sequence.
 ACCESSION B6027821
 VERSION B6027821.1 GI:12416827
 KEYWORDS EST.
 SOURCE human.
 ORGANISM human.

REFERENCE
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
 Plate: L1AM10077 row: 1 column: 19
 High quality sequence stop: 637.
 Location/Qualifiers

FEATURES

source

1..905
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4389594"
 /clone_lib="NIH_MGC_86"
 /tissue_type="Osteosarcoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: bone; Vector: pCMV-Sport6; Site.1: NotI;
 Site.2: SalI; Cloned unidirectionally; oligo-dt primed.
 Average insert size 1.53 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 BASE COUNT 280 a 148 c 221 g 256 t
 ORIGIN

Query Match 49.8%; Score 635; DB 11; Length 905;
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 635; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 TTCTAGGCAATTTGGCTATGATTTGACGAAGGGAACCGACTGGAATCTTGCATTA 566
 Db 1 TTCTAGGCAATTTGGCTATGATTTGACGAAGGGAACCGACTGGAATCTTGCATTA 60
 QY 567 ATGTGATGATCTGATGACGACCAATTAATGATATCAATGATGTCAACGCTGAAAC 626
 Db 61 ATGTGATGATCTGATGACGACCAATTAATGATATCAATGATGTCAACGCTGAAAC 120
 QY 627 CTGGCTACTTGAAGCTACTGTGACCTGTTTAAAGGTATTAAGGTCTGATGAAAC 686
 Db 121 CTGGCTACTTGAAGCTACTGTGACCTGTTTAAAGGTATTAAGGTCTGATGAAAC 180
 QY 687 CAGAAATGAGTTTGGCTTTATGACGAATTTAAAGATTAAGAGACTTGGCATGATATTA 746
 Db 181 CAGAAATGAGTTTGGCTTTATGACGAATTTAAAGATTAAGAGACTTGGCATGATATTA 240

OY	747	TTAAACACATCTATGACCACTTTGGAAAGCATTTAGTACTATTAAGAAAAACGATTTGAAAAAGGA	806
Db	241	TTAAAGACATCTATGACCACTTTGGAAAGCATTTAGTACTATTAAGAAAAACGATTTGAAAAAGGA	300
OY	807	TCAGTTGCATGAATATACAACTTTGTCTGAGAGGCCCTTCAGTGTGATCTGATGTCTGCCA	866
Db	301	TCAGTTGCATGAATATACAACTTTGTCTGAGAGGCCCTTCAGTGTGATCTGATGTCTGCCA	360
OY	867	GAGCCATTTTGGAATGCTTTACCAACCACCTGTGAAATCTGCTGTGCACAGTACCAACAGAGC	926
Db	361	GAGCCATTTTGGAATGCTTTACCAACCACCTGTGAAATCTGCTGTGCACAGTACCAACAGAGC	420
OY	927	TGCTAAGTGTGCTCCATCCACAGAAAACTATATGATTTCTGTGSAATCAAGCTGATA	986
Db	421	TGCTAAGTGTGCTCCATCCACAGAAAACTATATGATTTCTGTGSAATCAAGCTGATA	480
OY	987	TTCGTACATCGTTCATCTGAGTGTATTAGAAAGTAAAGTAGAGCTTTTCAAAGCTTT	1046
Db	481	TTCGTACATCGTTCATCTGAGTGTATTAGAAAGTAAAGTAGAGCTTTTCAAAGCTTT	540
OY	1047	AAATTTGTAGACTCATCTTAACATAAAGTAATTTGCTGTGACATATCCAAATATCTACG	1106
Db	541	AAATTTGTAGACTCATCTTAACATAAAGTAATTTGCTGTGACATATCCAAATATCTACG	600
OY	1107	AATGTTATCCATTAAGAATTTTTCATATCTCAA	1141.
Db	601	AATGTTATCCATTAAGAATTTTTCATATCTCAA	635

RESULT	6
BG574468	
LOCUS	
DEFINITION	BG574468 915 bp mRNA
ACCESSION	U025564.1 F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4705253
VERSION	BG574468 mRNA sequence.
KEYWORDS	BG574468.1 GI:13582121 EST
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	1 (bases 1 to 915)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: WGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM10571 row: j column: 06
High quality sequence stop: 830.

FEATURES	location/Qualifiers
source	1. ,915

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4705253"
/clone_lib="NHU_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/name="Organ breast; Vector: pCMV-Sport6; Site.1: NotI; Site.2: SalI; Cloned unidirectionally; oligo-dT primed
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies

```

BASE COUNT	269 a	185 c	231 g	230 t
ORIGIN				

Query Match	49.68;	Score 632;	DB 11;	Length 915;
Best Local Similarity	: 100.0%;	Pred. No. 0;		
Matches ,632;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

y 29 GTCTGGGCGCCGCTGCGGAGCTGTGGCTCTGTGGACAGCGCGCGGACGAGACTCCGGCACT 88
 |||||
 |||||
 b 1 GTCTGGGCGCCGCTGCGGAGCTGTGGCTCTGTGGACAGCGCGCGGACGAGACTCCGGCACT 60
 |||||
 y 89 ATGAGCGGCTTCAGCACCCGAGGAGCGCGCGCCCTTCCTCCCTGGAGTACCGAGTCTTC 148
 |||||
 |||||
 b 61 ATGAGCGGCTTCAGCACCCGAGGAGCGCGCGCCCTTCCTCCCTGGAGTACCGAGTCTTC 120
 |||||
 y 149 CTCGAAATGAGAAAGACAAATATATATCTCCATTTCAATGATATCCAAATTTATTCAGAT 208
 |||||
 |||||
 b 121 CTCGAAATGAGAAAGACAAATATATATCTCCATTTCAATGATATCCAAATTTATTCAGAT 180
 |||||
 y 209 AAGGATGTTTTCACATGCTAGTTGAAGTACCAAGCTGCTAATGCAGAAATGGAGATT 268
 |||||
 |||||
 b 181 AAGGATGTTTTCACATGCTAGTTGAAGTACCAAGCTGCTAATGCAGAAATGGAGATT 240

269 GCATAAAGACCCCTTTAAACCTTTAAACAGATGTGAAAAAGGAAAACTCGCAT 328
241 GCTACAAAGACCCCTTTAAACCTTTAAACAGATGTGAAAAAGGAAAACTTGCGCAT 300
329 GTTGGCAATTTGTCCTCGATATAAGGATATATCTGAACTATGTGCCATCCTCAGACT 388
301 GTTGGCAATTTGTCCTCGATATAAGGATATATCTGAACTATGTGCCATCCTCAGACT 360

389	TGGGAGACC	CCAGG	CAC	CAATG	ATAAC	TACTGG	CTGTTGG	TGTGAC	AATGACC	CAATT	448	
361	TGGGAAG	ACC	CCAGG	GGC	CAATG	ATAAC	TACTGG	CTGTTGG	TGTGAC	AATGACC	CAATT	420
449	GATGTGT	GGAATT	TGGAAG	CAAG	TATG	CTCAAG	AGGTG	GAATTA	ATTGG	CGTGAAG	TT	508
421	GATGTGT	GGAATT	TGGAAG	CAAG	TATG	CTCAAG	AGGTG	GAATTA	ATTGG	CGTGAAG	TT	480
509	CTAGG	CATATT	TGGCT	TATG	ATTG	ACCAAG	GGG	GAAC	CCG	ACTGGA	AGTATT	568
481	CTAGG	CATATT	TGGCT	TATG	ATTG	ACCAAG	GGG	GAAC	CCG	ACTGGA	AGTATT	540

y 569 GTGGATGATCCTCGATGCACGCAATTAATATGATATCAATGATGTCAAAAGGGTGAACCT 628
 |||||
 541 GTGGATGATCCTCGATGCACGCAATTAATATGATATCAATGATGTCAAAAGGGTGAACCT 600
 |||||
 y 629 GGGTACTAGAAAGCTACTGTGAGCTGGTTAG 660
 |||||

b	601	GCGCTACTGTAAGACGTACTGTGCAGCTGGTTTAA	632
RESULT	7		
BEB887465			
COCUS			
INTEGRATION	6015108111ff	NH_MGC_71 Homo sapiens cDNA clone IMAGE:3909506 5',	
CCESSTION	mRNA sequence:		
VERSION	BEB887465	628 bp mRNA EST 20-OCT-2000	
	BEB887465.1	GI:10342782	

KEYWORDS	EST
SOURCE	human.
ORGANISM	Homo sapiens.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 628)
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished (1999)
	Contact: Robert Strausberg, Ph.D.
	Email: cgabbs@remail.nih.gov
	Tissue Procurement: ATCC

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.

found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LIAM9723. row: f column: 03
 High quality sequence stop: 626.
 Location/Qualifiers

FEATURES

source

1. 628

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="3909506"
 /clone_lib="NH_MGC_71"
 /tissue_type="Leiomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-Sport6; Site: 1; Not:
 Site: 2; Salt: Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 kb."
 Average insert size 2.1 kb.
 Average insert size 2.1 kb.

BASE COUNT

203

a 107

c 151

g 167

t

Query Match

Best Local Similarity 49.3%; Score 628; DB 11; Length 628;
 Matches 628; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

395 GACCCGAGGACAAATGATTAACATAGCTGCTGTTGCTGACAAATGACCAATGATG 454
 1 GACCCGAGGACAAATGATTAACATAGCTGCTGTTGCTGACAAATGACCAATGATG 60
 455 TGTGAATTTGGAAGCAAGATATGTCAGAGAGTGAATTAATTTGGCTGAAGTTCTAG 514
 61 TGTGAATTTGGAAGCAAGATATGTCAGAGAGTGAATTAATTTGGCTGAAGTTCTAG 120
 515 ATATTGGCTATGATGTCAGAGAGGGAACCGACTGGAAGTCCATTATTTGATG 574
 121 ATATTGGCTATGATGTCAGAGAGGGAACCGACTGGAAGTCCATTATTTGATG 180
 575 GATCTGATGACCAATTAATTAATGATATGATGTCAGAGAGTGAAGTTCTAG 634
 181 GATCTGATGACCAATTAATTAATGATATGATGTCAGAGAGTGAAGTTCTAG 240
 635 TTGAAGGCTACGTGAGCTGCTTGAAGGATTAAGTTCCTGATGGAAGCAAGAAAT 694
 241 TTGAAGGCTACGTGAGCTGCTTGAAGGATTAAGTTCCTGATGGAAGCAAGAAAT 300
 695 GACTTTGCGTTTAATGAGAAATTAAGATAGAGACTTTGCCATTATTTAAAGC 754
 301 GACTTTGCGTTTAATGAGAAATTAAGATAGAGACTTTGCCATTATTTAAAGC 360
 755 ACTCATGACCAATTTAGAGATTAAGTACTAAGAAACGATGGAAGAAATCAAGTTGC 814
 361 ACTCATGACCAATTTAGAGATTAAGTACTAAGAAACGATGGAAGAAATCAAGTTGC 420
 815 ATGAATGCACTTTGCTGAGAGCCCTTAAGTGTGATGCTGATGCTGCAAGCCATT 874
 421 ATGAATGCACTTTGCTGAGAGCCCTTAAGTGTGATGCTGATGCTGCAAGCCATT 480
 875 GTGATGCTTTACACACCCCTTGATTCGCTGACAGATGACCAACAGAGTGATAG 934
 481 GTGATGCTTTACACACCCCTTGATTCGCTGACAGATGACCAACAGAGTGATAG 540
 935 TGTGTCATACAGAAAACCTAATGATTTCTCTGGAATTAAGAGTATTTGCTACA 994
 541 TGTGTCATACAGAAAACCTAATGATTTCTCTGGAATTAAGAGTATTTGCTACA 600
 995 TGTGTCATACAGAAAACCTAATGATTTCTCTGGAATTAAGAGTATTTGCTACA 1022
 601 TGTGTCATACAGAAAACCTAATGATTTCTCTGGAATTAAGAGTATTTGCTACA 628

RESULT

8

LOCUS

AM009649

660 bp

mrna

EST

08-MAR-2000

DEFINITION

ws85603.x1 NCI CGAP Co3 Homo sapiens CDNA clone image:2504764.3'

similar to SW:IPYR_BOVIN P37980 INORGANIC PHOSPHATASE ;, mRNA

sequence.
 AM009649
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

human.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

sequence.
 AM009649
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgsb@bbs-remail.nih.gov
 Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.,
 Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arraying: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/dbp/image/image.html
 Insert Length: 418 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 457.
 Location/Qualifiers

FEATURES

source

1. 660

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="2504764"
 /clone_lib="NCI CGAP Co3"
 /sex="pooled"
 /tissue_type="colon"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was prepared from 12 pooled bulk tumor samples and primed
 with a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the
 modified pT73 vector. Library went through one round of
 normalization."
 Location/Qualifiers

BASE COUNT

197

a 129

c 110

g 221

t 3

others

Query Match

Best Local Similarity 49.0%; Score 625; DB 10; Length 660;
 Matches 625; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

651 ACTGTTTGAAGGATTAAGTCTGATGGAACCAAGAAATGATTTGGCTTAATG 710
 628 ACTGTTTGAAGGATTAAGTCTGATGGAACCAAGAAATGATTTGGCTTAATG 569
 711 CAGATTTAAGTAAGACTTTGCCATTGATTTTAAAGCACTATGACATGGA 770
 568 CAGATTTAAGTAAGACTTTGCCATTGATTTTAAAGCACTATGACATGGA 509
 771 AAGCATTAGGATTAAGAAAGCAATGGAAGGATGATGATGATTAACATTGT 830
 508 AAGCATTAGGATTAAGAAAGCAATGGAAGGATGATGATGATTAACATTGT 449
 831 CTGAGAGCCCTTCAAGTGTGATCTGATGCTGCCAGAGCCATTGATGATGCTTACAC 890
 448 CTGAGAGCCCTTCAAGTGTGATCTGATGCTGCCAGAGCCATTGATGATGCTTACAC 389
 891 CACCTGTGATGCTGCTGACAGTCAAGAGAGAGCTGATAGAGTTCATACACAGA 950
 388 CACCTGTGATGCTGCTGACAGTCAAGAGAGAGCTGATAGAGTTCATACACAGA 329
 951 AAACCTAAGAGATTTCTGGAATCAAGCTGATATTTCTCATCATGTTTCATCTGAT 1010

Db 328 AAAAATAATGAGATTTCTCTGGAATACAGCTGATATTGCTACATCGTTCATCTGGAT 269
QY 1011 GATATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1070
Db 268 GATATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 209
QY 1071 AAGTAAATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1130
Db 208 AAGTAAATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 149
QY 1131 TCATATCCACTAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1190
Db 148 TCATATCCACTAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 89
QY 1191 GAAGTCACTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1250
Db 88 GAAGTCACTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 29
QY 1251 GTTAGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1275
Db 28 GTTAGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4

RESULT 9
Bg113976 772 bp mRNA EST 08-MAY-2001
LOCUS 602674374F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4796875 5',
DEFINITION mRNA sequence.
Bg113976
ACCESSION Bg113976.1 GI:13992907
VERSION Bg113976.1 GI:13992907
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 772)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs.fremail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LLM10682 row: C column: 20
High quality sequence stop: 767.
Location/Qualifiers
1. 772
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4796875"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript (modified
pBluescript KS+); Site: 1: BamHI; Site 2: SalI-XhoI (glucag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to ROP 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 235 a 146 c 197 g 194 t
ORIGIN

Query Match 48.5%, Score 618; DB 11; Length 772;

Best Local Similarity: 100.0%; Pred. No. 4e-308;
Matches 618; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 79 CTCGGACATGAGAGGCGCTTCAGCAGCAGAGAGGCGCGCCCTTCCTCCAGAGTA 138
Db 57 CTCGGACATGAGAGGCGCTTCAGCAGCAGAGAGGCGCGCCCTTCCTCCAGAGTA 116
QY 139 CCGAGCTCTCTCAAAAAATGAGAAAGGACAAATATATCTCCATTTGATATATTCAT 198
Db 117 CCGAGCTCTCTCAAAAAATGAGAAAGGACAAATATATCTCCATTTGATATATTCAT 176
QY 199 TTATGACATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 258
Db 177 TTATGACATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 236
QY 259 AATGAGATTTGCTACAAAGAGCCCTTTAAACCTTTAAACAGATGTAAGAAAGGAA 318
Db 237 AATGAGATTTGCTACAAAGAGCCCTTTAAACCTTTAAACAGATGTAAGAAAGGAA 296
QY 319 ACTTGCTATGTTGGCAATTTGTTCCGATTAAGATATATCTGAACTATGTCAT 378
Db 297 ACTTGCTATGTTGGCAATTTGTTCCGATTAAGATATATCTGAACTATGTCAT 356
QY 379 CCTCAGACTTGGGAGAGCCAGGCGACATGATTAACATCTGCTGTTGTGACAA 438
Db 357 CCTCAGACTTGGGAGAGCCAGGCGACATGATTAACATCTGCTGTTGTGACAA 416
QY 439 TGACCCAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 498
Db 417 TGACCCAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 476
QY 499 CGTGAAGTTCTAGCATATTTGGCTATGATGATGATGATGATGATGATGATGATGATGAT 558
Db 477 CGTGAAGTTCTAGCATATTTGGCTATGATGATGATGATGATGATGATGATGATGATGAT 536
QY 559 TGCCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 618
Db 537 TGCCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 596
QY 619 GCTGAACCTGGCTACTAGAAAGCTATGATGATGATGATGATGATGATGATGATGATGAT 678
Db 597 GCTGAACCTGGCTACTAGAAAGCTATGATGATGATGATGATGATGATGATGATGATGAT 656
QY 679 TGGAAACACAGAAATGA 696
Db 657 TGGAAACACAGAAATGA 674

RESULT 10
BF791373 955 bp mRNA EST 12-JAN-2001
LOCUS 602251332P1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4344030 5',
DEFINITION mRNA sequence.
BF791373
ACCESSION BF791373.1 GI:12096427
VERSION BF791373.1 GI:12096427
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 955)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs.fremail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov

Y	534	AAGGGAAACCGACGCGSAAAGCATTCGTCATTAAATGSGATGATCCTGATGAGCAANT	593
Y	534	AAGGGAAACCGACGCGSAAAGCATTCGTCATTAAATGSGATGATCCTGATGAGCAANT	593
Db	121	AAGGGAAACCGACGCGSAAAGCATTCGTCATTAAATGSGATGATCCTGATGAGCAANT	180
Y	594	ATAATGATATCATGATATCAAAAGCGCTGAAAACCTGGCTACTTAGAACCTACTGTGACT	653
Db	181	ATAATGATATCATGATATCAAAAGCGCTGAAAACCTGGCTACTTAGAACCTACTGTGACT	240
Y	654	GCTTTAGAGGTATATAGGTTCTGTATGAGAAAAACCGAAAATGATGATTTGCGTTAATGAG	713
Db	241	GCTTTAGAGGTATATAGGTTCTGTATGAGAAAAACCGAAAATGATGATTTGCGTTAATGAG	300
Y	714	AATTTAAGATATAGACCTTGGCCATTTGATATTTATTAAGAACACATCATGACCTTGGGAAG	773
Db	301	AATTTAAGATATAGACCTTGGCCATTTGATATTTATTAAGAACACATCATGACCTTGGGAAG	360
Y	774	CATTAGTGACATAGAAAAAGAAATGGAAGAAATGATGATGATGATGATGATGATGATG	833
Db	361	CATTAGTGACATAGAAAAAGAAATGGAAGAAATGATGATGATGATGATGATGATGATG	420
Y	834	AGAGCCCTTTAAGTGTATCTCTATGCTGTCAGAGCCATTTGATGATGATGATGATGATG	893
Db	421	AGAGCCCTTTAAGTGTATCTCTATGCTGTCAGAGCCATTTGATGATGATGATGATGATG	480
Y	894	CCTGTGATTCGCTGCGACAGTACACAGACAGAGCTGATTAAGGTTTCATCCACGAAAA	953
Db	481	CCTGTGATTCGCTGCGACAGTACACAGACAGAGCTGATTAAGGTTTCATCCACGAAAA	540
Y	954	ACTAATGAGATTTCTCTGGAATACAAAGCTGATTTGCTACATCGTTTCATCTGG	1008
Db	541	ACTAATGAGATTTCTCTGGAATACAAAGCTGATTTGCTACATCGTTTCATCTGG	595
RESULT	13		
LOCUS	BF966851/c		
DEFINITION	60228650371 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4375297 3', mRNA sequence.		
ACCESSION	BF966851		
VERSION	BF966851.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 998)		
TITLE	NIH-MGC http://mgi.nci.nih.gov/		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: rgs@nih.gov		
	Tissue Procurement: Miklos Palkovits, M.D., Ph.D.		
	cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki		
	Toshiyuki and Piero Carninci (RIKEN)		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://Image.llnl.gov		
	Plate: LAM10040		
	row: f		
	column: 02		
	High quality sequence start: 52		
	High quality sequence stop: 761.		
	Location/Qualifiers		
	1. 998		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4375297"		
	/clone_lib="NIH MGC 95"		
	/tissue_type="hippocampus"		
	/lab_host="DH10B"		
	/note="Organ: Brain; Vector: Bluescript (modified		
	bluescript KS+); Site:1: BamHI; Site:2: SalI-XhoI (gtcga		

); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.5 kb and normalized to R0.5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NIGRI, National Institutes of Health). Note: this is a NIH-MGC Library."

BASE COUNT 249 a 245 c 182 g 322 t
ORIGIN

Query Match 46.4% Score 592; DB 11; Length 998;
Best Local Similarity 100.0%; Pred. No. 1.1e-294;
Matches 592; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY AACGGCTGAACCTGGCTACTTAGAAGCTACGTGCTGCTTTAGAAAGTAAAGTTC 674
DB AACGGCTGAACCTGGCTACTTAGAAGCTACGTGCTGCTTTAGAAAGTAAAGTTC 613
QY CTGATGAGAAACCCAGAAATGAGTTGCGTTTAATGCAAAATTTAAAGTAAGACTTTG 734
DB CTGATGAGAAACCCAGAAATGAGTTGCGTTTAATGCAAAATTTAAAGTAAGACTTTG 553
QY CCATTGATATTTTAAAGCACTCATGACCAATTGGAAGCAATTAGACTAAGAAACGA 794
DB CCATTGATATTTTAAAGCACTCATGACCAATTGGAAGCAATTAGACTAAGAAACGA 493
QY ATGGAAGAAAGCAATGCTGATGAAATGCAATTTGCTGAGAGCCCTTCAAGTGTGAGC 854
DB ATGGAAGAAAGCAATGCTGATGAAATGCAATTTGCTGAGAGCCCTTCAAGTGTGAGC 433
QY CTGATGCTCCAGAGCACTTGTGAGTGTACCAACCCCTGTGATGCTGCTGACAG 914
DB CTGATGCTCCAGAGCACTTGTGAGTGTACCAACCCCTGTGATGCTGCTGACAG 373
QY TACCAACAGAGCTGATAGTGTGCTCATCCACAGAAATTAAGATGATTTCTGGA 974
DB TACCAACAGAGCTGATAGTGTGCTCATCCACAGAAATTAAGATGATTTCTGGA 313
QY TACCAAGCTATTTGCTGATGCTGCTCATCTGAGATGTTGAAGTAAAGTAGAGCT 1034
DB TACCAAGCTATTTGCTGATGCTGCTCATCTGAGATGTTGAAGTAAAGTAGAGCT 253
QY TTTCAAGCTTTAAATTTAGAACCTCATTAAGTAAATTTCTGCTGACTATC 1094
DB TTTCAAGCTTTAAATTTAGAACCTCATTAAGTAAATTTCTGCTGACTATC 193
QY CAATATCTCAGAAATGTTATCCATCTAAGCAATTTTCAATCTCAACTAAGTAACTTT 1154
DB CAATATCTCAGAAATGTTATCCATCTAAGCAATTTTCAATCTCAACTAAGTAACTTT 133
QY TTAGCAATGCTTAAATATCAAGCAAGTGTGCTTTGGAAGTCACTTGTGTAAT 1206
DB TTAGCAATGCTTAAATATCAAGCAAGTGTGCTTTGGAAGTCACTTGTGTAAT 81

RESULT 14

LOCUS BG676397 932 bp mRNA EST 01-MAY-2001
DEFINITION 602622839P1 NCL_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4747977 5',
mRNA sequence.

ACCESSION BG676397
VERSION BG676397.1 GI:13907793
KEYWORDS EST.

SOURCE

human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 932)
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM10598 row: n column: 10
High quality sequence stop: 837.

FEATURES

source
location/qualifiers
1..932
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4747977"
/clone_1ib="NCL_CGAP_Skn4"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (TL phage-resistant)"
/note="Organ: skin; Vector: pcMV-SPOrt6; Site_1: NCL; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCL_CGAP Library."
BASE COUNT 309 a 155 c 208 g 260 t
ORIGIN

Query Match 46.0% Score 586; DB 11; Length 932;
Best Local Similarity 99.9%; Pred. No. 1.4e-291;
Matches 706; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY CTACAAAGAGCCCTTTAAACCTATTAAACAAGATGTAAGAAAGAAACCTGCTATG 329
DB CTACAAAGAGCCCTTTAAACCTATTAAACAAGATGTAAGAAAGAAACCTGCTATG 60
QY TTGCAATTTGTTCCCGTATTAAGATATATGCACTATGTCCTGACTT 389
DB TTGCAATTTGTTCCCGTATTAAGATATATGCACTATGTCCTGACTT 120
QY TGGCAATTTGTTCCCGTATTAAGATATATGCACTATGTCCTGACTT 120
DB TGGCAATTTGTTCCCGTATTAAGATATATGCACTATGTCCTGACTT 120
QY GGGAGAGCCAGGCGACATGATTAACATGCTGCTGTTGCTGACATGACCAATG 449
DB GGGAGAGCCAGGCGACATGATTAACATGCTGCTGTTGCTGACATGACCAATG 179
QY ATGTGTGTAATTTGAAGCAAGTATGCAAGAGGTGAATTAATTTGCGTGAAGTTC 509
DB ATGTGTGTAATTTGAAGCAAGTATGCAAGAGGTGAATTAATTTGCGTGAAGTTC 239
QY ATGTGTGTAATTTGAAGCAAGTATGCAAGAGGTGAATTAATTTGCGTGAAGTTC 239
DB ATGTGTGTAATTTGAAGCAAGTATGCAAGAGGTGAATTAATTTGCGTGAAGTTC 239
QY TTAGCATATTTGCTATGATGTCGAAGGGAACCCGACTGGAATCTGTCCTATATG 569
DB TTAGCATATTTGCTATGATGTCGAAGGGAACCCGACTGGAATCTGTCCTATATG 299
QY TAGGATATTTGCTATGATGTCGAAGGGAACCCGACTGGAATCTGTCCTATATG 299
DB TAGGATATTTGCTATGATGTCGAAGGGAACCCGACTGGAATCTGTCCTATATG 299
QY TGGATGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 629
DB TGGATGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 359
QY GCTACTTGAAGCTACTGCTGAGCTGTTAGAGGTATTAAGTCTGATGGAAGAACAG 419
DB GCTACTTGAAGCTACTGCTGAGCTGTTAGAGGTATTAAGTCTGATGGAAGAACAG 419
QY AAAATGAGTTTGGCTTTAATGAGAAATTTAAAGATTAAGAGCTTTCCATTTGATATTA 479
DB AAAATGAGTTTGGCTTTAATGAGAAATTTAAAGATTAAGAGCTTTCCATTTGATATTA 479
QY AAAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 809
DB AAAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 539
QY GTTGCATGAATTAACAATTTGCTGAGAGCCCTTCAAGTGTGATGCTGACAGAG 869
DB GTTGCATGAATTAACAATTTGCTGAGAGCCCTTCAAGTGTGATGCTGACAGAG 869
QY GTTGCATGAATTAACAATTTGCTGAGAGCCCTTCAAGTGTGATGCTGACAGAG 599
DB GTTGCATGAATTAACAATTTGCTGAGAGCCCTTCAAGTGTGATGCTGACAGAG 599
QY CCATTGTGATGCTTTTACCAACCCCTGTGATGCTGCAACATCAACAGACGTGG 929
DB CCATTGTGATGCTTTTACCAACCCCTGTGATGCTGCAACATCAACAGACGTGG 659

Best Local Similarity 100.0%; Pred. No. 5,4e-289;
Matches 581; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

39 CGTGGCGGCTGGTGGCTGTCTGTGGACAGCGCGCGGAGACATCCGGCACTATGAGGGCT 98
1 CGTGGCGGCTGGTGGCTGTCTGTGGACAGCGCGCGGAGACATCCGGCACTATGAGGGCT 60
99 TCAGCAGCAGGAGCGCGCGCGCGCTTCTCCCTGGAGTACAGAGTCTCCCTCAAAATG 158
61 TCAGCAGCAGGAGCGCGCGCGCGCTTCTCCCTGGAGTACAGAGTCTCCCTCAAAATG 120
159 AGAAGAGCAAT 218
121 AGAAGAGCAAT 180
219 TTCACATGTAGTGTGAAGTACACAGCTGTGTATATGCAAAATGAGATTGTACAAAGG 278
181 TTCACATGTAGTGTGAAGTACACAGCTGTGTATATGCAAAATGAGATTGTACAAAGG 240
279 ACCCTTAAACCTTATTAACAAGATGTGAAAAAGAAACCTTCCCTATGTTCGAAT 338
241 ACCCTTAAACCTTATTAACAAGATGTGAAAAAGAAACCTTCCCTATGTTCGAAT 300
339 TGTTCCTGATATAAGGATATATCTGGAGTATGTGCAATCCCTCAGACTTGGGAAGAC 398
301 TGTTCCTGATATAAGGATATATCTGGAGTATGTGCAATCCCTCAGACTTGGGAAGAC 360
399 CAGGCAAT 458
361 CAGGCAAT 420
459 AATTTGGAAGCAAGTATGTGCAAGAGGTGAATTAATTTGGCGTGAAGTTCAGCATAT 518
421 AATTTGGAAGCAAGTATGTGCAAGAGGTGAATTAATTTGGCGTGAAGTTCAGCATAT 480
519 TGGCTATGATTTGACGAAGGGAACCGACTGGAAGTCAATTCCTATTAATGTGATGATC 578
481 TGGCTATGATTTGACGAAGGGAACCGACTGGAAGTCAATTCCTATTAATGTGATGATC 540
579 CTGATGAGGCAATTAAT 619
541 CTGATGAGGCAATTAAT 581

RESULT 17
BF966192 882 bp mRNA EST 23-JAN-2001
LOCUS 602286503F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4375297 5'
DEFINITION mRNA sequence.
ACCESSION BF966192
VERSION BF966192.1 GI:12333407
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 882)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-t@mail.nih.gov
Tissue: Procurement: Miklos Palokovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHAM10040 row: f column: 02
High quality sequence score: 638.
Location/Qualifiers

source
1. 882
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4375297"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.5 kb and
normalized for RFLP 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC library."

BASE COUNT 236 a 187 c 240 g 219 t
ORIGIN

Query Match 45.4%; Score 579; DB 11; Length 882;
Best Local Similarity 100.0%; Pred. No. 5,8e-288;
Matches 579; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

23 TTGTCAATGCGCGCGCGCTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 82
8 TTGTCAATGCGCGCGCGCTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 67
83 GGCATATGAGGCGGCTTACAGCAGGAGCGCGCGCGCTTCTCCCTGAGTACGA 142
68 GGCATATGAGGCGGCTTACAGCAGGAGCGCGCGCGCTTCTCCCTGAGTACGA 127
143 GTCTTCTCAAAAATGAGAAAGAGCAATATATATATATATATATATATATATATATAT 202
128 GTCTTCTCAAAAATGAGAAAGAGCAATATATATATATATATATATATATATATATAT 187
203 GCAGATAGAGGTGTTCACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 262
188 GCAGATAGAGGTGTTCACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 247
263 GAGATGTGTACAAAGACCTTTAAACCTTATTAACAAGATATGTGAAAGAAACCTT 322
248 GAGATGTGTACAAAGACCTTTAAACCTTATTAACAAGATATGTGAAAGAAACCTT 307
323 CCTATGTGTGCAATTTGTTCCTGATTAAGAGATATCTGGAATATGTGCAATTCCT 382
308 CCTATGTGTGCAATTTGTTCCTGATTAAGAGATATCTGGAATATGTGCAATTCCT 367
383 CAGACTTGGGAGACCCAGGAGGACAAATGATAAATGCTGTGTGTGTGTGTGTGTGTGT 442
368 CAGACTTGGGAGACCCAGGAGGACAAATGATAAATGCTGTGTGTGTGTGTGTGTGTGT 427
443 CCAATTTGATGTGTGAAATTTGGAAGCAAGTATGTGCAAGAGTGAATAATTTGGCGTG 502
428 CCAATTTGATGTGTGAAATTTGGAAGCAAGTATGTGCAAGAGTGAATAATTTGGCGTG 487
503 AAAGTTCTAGCATATTTGCTATGATTAACGAAGGGAACCGCATGTGAAAGTCTTCC 562
488 AAAGTTCTAGCATATTTGCTATGATTAACGAAGGGAACCGCATGTGAAAGTCTTCC 547
563 AATTAATGTGATGATTCCTATGATGACGACCAATTAATGAT 601
548 AATTAATGTGATGATTCCTATGATGACGACCAATTAATGAT 586

RESULT 18
BG777143 711 bp mRNA EST 15-MAY-2001
LOCUS BG777143
DEFINITION 602664359F1 NIH_MGC_59 Homo sapiens cDNA clone IMAGE:4809591 5',
mRNA sequence.
ACCESSION BG777143
VERSION BG777143.1 GI:14047460
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 711)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE Unpublished (1999)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov
Plate: LNCM1664 row: e column: 16
High quality sequence stop: 700.
Location/Qualifiers
1. 711
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4809591"
/clone_lib="NIH-MGC_59"
/tissue_type="mucoepidermoid carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Lung; Vector: pDNR-LIB (Clontech); Site_1: SfiI (99cgcgcgcgcgc); Site_2: SfiI (99cgcgcgcgcgc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCGCATATGAGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGCGCATG-dt(30)BN-3' (where B = A, C, G, or T). Average insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 228 a 122 c 171 g 190 t
ORIGIN

Query Match 45.3%; Score 577; DB 11; Length 711;
Best Local Similarity 100.0%; Pred. No. 6.1e-287;
Matches 577; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 AGGAAACCTTCGCTATGTTGCGAATTTGTCCTGATTAAGGATATATCTGGAATATG 372
OY 313 AGGAAACCTTCGCTATGTTGCGAATTTGTCCTGATTAAGGATATATCTGGAATATG 372
Db 1 AGGAAACCTTCGCTATGTTGCGAATTTGTCCTGATTAAGGATATATCTGGAATATG 60
OY 373 TGGCATTCCCTCAGACTTGGGAAGACCCAGGACACATATAACTACTGCTGTGG 432
Db 61 TGGCATTCCCTCAGACTTGGGAAGACCCAGGACACATATAACTACTGCTGTGG 120
OY 433 TGACATATACCAATTTGATGTGTGAATTTGAAGCAAGGATATGCAAGGATGAAT 492
Db 121 TGACATATACCAATTTGATGTGTGAATTTGAAGCAAGGATATGCAAGGATGAAT 180
OY 493 AATTGGCGTGAAGTCTAGGCAATTTGCGTATGATTTGACGAAGGGAACCGATGGA 552
Db 181 AATTGGCGTGAAGTCTAGGCAATTTGCGTATGATTTGACGAAGGGAACCGATGGA 240
OY 553 AGTCTATCCCATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 612
Db 241 AGTCTATCCCATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
OY 613 CAACGGCTGAAGCTGCTACTTAGAAGCTACTGAGACGCTGTTAGAGGATTAAGGT 672
Db 301 CAACGGCTGAAGCTGCTACTTAGAAGCTACTGAGACGCTGTTAGAGGATTAAGGT 360
OY 673 TCTGATGGAAGCAAGAAATGAGTTGCGTTTAATGACGAATTTAAAGATAGAGCTT 732

Db 361 TCCGTATGGAAGAACGAGAAATGAGTTGCTTTAATGCAGATTTAAAGATAGGACTT 420
OY 733 TCCCATATGATTTATTAAGACACCTATGACATTTGGAAGGATATGATTAAGAAAC 792
Db 421 TCCCATATGATTTATTAAGACACCTATGACATTTGGAAGGATATGATTAAGAAAC 480
OY 793 GAATGGAAGAAATCATGATTTGATGATTAAGCAATTTGCTGAGAGCCCTTCAAGTGA 852
Db 481 GAATGGAAGAAATCATGATTTGATGATTAAGCAATTTGCTGAGAGCCCTTCAAGTGA 540
OY 853 TCCGTATGCTGCAGACGACCATTTGTGATGCTTTACCA 889
Db 541 TCCGTATGCTGCAGACGACCATTTGTGATGCTTTACCA 577

RESULT 19
LOCUS AM262869/c 575 bp mRNA EST 28-DEC-1999
DEFINITION xg96e02.x1 NCI_CGAP_Brn53 Homo sapiens cDNA clone IMAGE:2758490 3' similar to SW:IPYR_BOVIN P37980 INORGANIC PYROPHOSPHATASE, mRNA sequence.

ACCESSION AM262869
VERSION AM262869.1 GI:6639685
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 575)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BRGP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Chris Moskalkuk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D., CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: www.bio.llnl.gov/bbrp/image/image.html

BASE COUNT 175 a 109 c 99 g 192 t
ORIGIN

Query Match 45.1%; Score 575; DB 10; Length 575;
Best Local Similarity 100.0%; Pred. No. 6.5e-286;
Matches 575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 701 GCGTTAATGCAAGATTTAAAGACCTTGGCATGATATTAAGACACTGAT 760
Db 575 GCGTTAATGCAAGATTTAAAGACCTTGGCATGATATTAAGACACTGAT 516
OY 761 GACCATTTGGAAGCAATTTAGTGAAGAAACGATGGAAGGATGATGATGATGAT 820


```

Db      515 GACCATTGGAAAGCATTAGTACTAAGAAAAAGAAATGAAAAAGAAATCAGTGCATGAAAT 456
QY      821 ACAACTTTGTCTGAGAGCCCTTCAAGTGTGATCTGATCTGCTCCAGAGCCATTGTGAT 880
Db      455 ACAACTTTGTCTGAGAGCCCTTCAAGTGTGATCTGATCTGCTCCAGAGCCATTGTGAT 396
QY      881 GCTTTTACCAACCCCTGTGATTTGCTGACAGTACCAACAGACGTGTGATAGTGGTTG 940
Db      395 GCTTTTACCAACCCCTGTGATTTGCTGACAGTACCAACAGACGTGTGATAGTGGTTG 336
QY      941 CACACACAGAAAAAGAAATGAGATTTCTGCGAATACAGCTATATGTGATCATGTGT 1000
Db      335 CACACACAGAAAAAGAAATGAGATTTCTGCGAATACAGCTATATGTGATCATGTGT 276
QY      1001 TCATCGAGTGTATGAGATTAAGTAGAGTTCCTTCAAGGCTTAAATTTGTAGAACT 1060
Db      275 TCATCGAGTGTATGAGATTAAGTAGAGTTCCTTCAAGGCTTAAATTTGTAGAACT 216
QY      1061 CATCTACATTAAGTAAATTTCTGCTGTGACTAATCCAAATATCTCAGAAATGTTATCCATCT 1120
Db      215 CATCTACATTAAGTAAATTTCTGCTGTGACTAATCCAAATATCTCAGAAATGTTATCCATCT 156
QY      1121 AAAGCAATTTTCAATATCTGCTGTGACTAATCCAAATATCTCAGAAATGTTATCCATCT 1180
Db      155 AAAGCAATTTTCAATATCTGCTGTGACTAATCCAAATATCTCAGAAATGTTATCCATCT 96
QY      1181 TTGTCAATTTGAAGTCACTTTGTAATAGATGTGCAAGGAGACATATTTGATGTATAT 1240
Db      95 TTGTCAATTTGAAGTCACTTTGTAATAGATGTGCAAGGAGACATATTTGATGTATAT 36
QY      1241 GTTACCATATGTTAGAAATTAATTTTGTGCTG 1275
Db      35 GTTACCATATGTTAGAAATTAATTTTGTGCTG 1

RESULT 20
Bg702955 714 bp mRNA EST 07-MAY-2001
LOCUS 602684915f1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4817474 5'
DEFINITION mRNA sequence.
ACCESSION Bg702955
VERSION Bg702955.1 GI:13974812
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 714)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs.fda.gov
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshitaki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL0717 row: n column: 03
High quality sequence stop: 711.
Location/Qualifiers
1. 714
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4817474"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pBluescript (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag

```

```

); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.5 kb and
normalized to R0.5. This is a primary library enriched
for full-length clones and constructed using the
cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH-MGC Library."
BASE COUNT 210 a 136 c 187 g 181 t
ORIGIN
Query Match 44.9%; Score 572; DB 11; Length 714;
Best Local Similarity 100.0%; Pred. No. 2.3e-284;
Matches 572; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 17 CTCCTCTTGTCAATGCGGCGCGGCTGCGTGTGTCGAGCGGCGGCGAG 76
Db 6 CTCCTCTTGTCAATGCGGCGCGGCTGCGTGTGTCGAGCGGCGGCGAG 65
QY 77 GACTCCGCGACATATGAGCGGCTTCAACAGAGAGCGGCGGCGGCTTCCCTGAG 136
Db 66 GACTCCGCGACATATGAGCGGCTTCAACAGAGAGCGGCGGCGGCTTCCCTGAG 125
QY 137 TACCGAGTCTTCTCTCAAAATGAGAAAGACATATATATTCATTCATGATATTC 196
Db 126 TACCGAGTCTTCTCTCAAAATGAGAAAGACATATATATTCATTCATGATATTC 185
QY 197 ATTATGCAATATGAGATGTGTTTCAATGATGATGAGTACAGCGTGTCTAATGCA 256
Db 186 ATTATGCAATATGAGATGTGTTTCAATGATGATGAGTACAGCGTGTCTAATGCA 245
QY 257 AAATGAGATATGCTCAAGAGACCCCTTAAACCTATTAACAGATGTGAAAAAGCA 316
Db 246 AAATGAGATATGCTCAAGAGACCCCTTAAACCTATTAACAGATGTGAAAAAGCA 305
QY 317 AAATGAGATATGCTCAAGAGACCCCTTAAACCTATTAACAGATGTGAAAAAGCA 376
Db 306 AAATGAGATATGCTCAAGAGACCCCTTAAACCTATTAACAGATGTGAAAAAGCA 365
QY 377 ATCCCTGACAGTGTGAGAGACCCAGGCGCAATGATTAACATGATGCTGTGTGAGC 436
Db 366 ATCCCTGACAGTGTGAGAGACCCAGGCGCAATGATTAACATGATGCTGTGTGAGC 425
QY 437 AATGACCAATGATGTGTGTAATTTGAGAGAGATGTGCAAGAGTGAATTAAT 496
Db 426 AATGACCAATGATGTGTGTAATTTGAGAGAGATGTGCAAGAGTGAATTAAT 485
QY 497 GCGGTGAAGTCTTGAAGCATATTTGCTATGATGATGAGAGAGAGAGAGAGAGC 556
Db 486 GCGGTGAAGTCTTGAAGCATATTTGCTATGATGATGAGAGAGAGAGAGAGAGC 545
QY 557 ATTGCATTAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 588
Db 546 ATTGCATTAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 577

RESULT 21
Bg702317 769 bp mRNA EST 07-MAY-2001
LOCUS 60268363f1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4816311 5'
DEFINITION mRNA sequence.
ACCESSION Bg702317
VERSION Bg702317.1 GI:13973536
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 769)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.

```

Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM10714 row: m column: 16
High quality sequence stop: 764.

FEATURES
source

1. 769
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4816311"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript (modified
pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcgag
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTTAA-3',
size-selected for average insert size 2.5 kb and
normalized to 10⁵. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT
ORIGIN

234 a 140 c 201 g 194 t

Query Match 44.7%; Score 570; DB 11; Length 769;
Best Local Similarity 100.0%; Pred. No. 2.6e-283;
Matches 570; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

79 CTCGGCAGTATGAGCGGCTTACAGACCGAGAGCGCGCCCTTCCCTCGAGTA 138
|||||
57 CTCGGCAGTATGAGCGGCTTACAGACCGAGAGCGCGCCCTTCCCTCGAGTA 116
|||||
139 CCGAGTCTTCTCAAAAATGAGAAAGACAATATATCTCCATTTATGATTCAT 198
|||||
117 CCGAGTCTTCTCAAAAATGAGAAAGACAATATATCTCCATTTATGATTCAT 176
|||||
199 TTATGAGATAGAGATGTGTTACATAGTAGTAGTACCGAGTGTCTATATGAAA 258
|||||
177 TTATGAGATAGAGATGTGTTACATAGTAGTAGTACCGAGTGTCTATATGAAA 236
|||||
259 AATGAGATGTGTAACAAGACCCCTTAACCCCTTAATTAACAGATGTGAAAAAGAAA 318
|||||
237 AATGAGATGTGTAACAAGACCCCTTAACCCCTTAATTAACAGATGTGAAAAAGAAA 296
|||||
319 ACTTGGTATGTTGCGAATTTGTTCCGATTAAGATATATCTGAACTATGTTGCAT 378
|||||
297 ACTTGGTATGTTGCGAATTTGTTCCGATTAAGATATATCTGAACTATGTTGCAT 356
|||||
379 CCTCGAGCTTGGAGAGACCCAGGCGCATGATTAACATACGTGCTTTGTGTGCAAA 438
|||||
357 CCTCGAGCTTGGAGAGACCCAGGCGCATGATTAACATACGTGCTTTGTGTGCAAA 416
|||||
439 TGACCCAAATGATGTGTGAATAATGGAACAAGGTATGTGCAAGAGGTGAATAATG 498
|||||
417 TGACCCAAATGATGTGTGAATAATGGAACAAGGTATGTGCAAGAGGTGAATAATG 476
|||||
499 CGTGAAGTCTAGGATATTTGGCTATGATTTGACGAAGGGGAAACGACGTGAAAGTAT 558
|||||
477 CGTGAAGTCTAGGATATTTGGCTATGATTTGACGAAGGGGAAACGACGTGAAAGTAT 536
|||||
559 TGCCATTAATGTGATGATCTGATGACGACCAATTAATATGATATCAATGATGCAAG 618
|||||
537 TGCCATTAATGTGATGATCTGATGACGACCAATTAATATGATATCAATGATGCAAG 596
|||||
619 GGTGAACCTGCTACTTAGAAGCTACTGT 648

Db 597 GCTGAACCTGCTACTTAGAAGCTACTGT 626

RESULT 22

LOCUS BG180287 567 bp mRNA EST 06-FEB-2001
DEFINITION 602331122F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4432636 5',
rna sequence
ACCESSION BG180287
VERSION BG180287.1 GI:12686990
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 567)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov

Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM10189 row: k column: 05
High quality sequence stop: 565.

FEATURES
source

1. 567
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4432636"
/clone_lib="NIH_MGC_91"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: prostate; Vector: pCMV-Sport6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 1.4 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT
ORIGIN

175 a 110 c 142 g 140 t

Query Match 44.5%; Score 567; DB 11; Length 567;
Best Local Similarity 100.0%; Pred. No. 8.8e-282;
Matches 567; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

60 GGCAGCGGCGGCGGAGACTCCGGCACTATGAGCGGCTTACGACCGAGAGCGCGCG 119
|||||
1 GGCAGCGGCGGCGGAGACTCCGGCACTATGAGCGGCTTACGACCGAGAGCGCGCG 60
|||||
120 GCGCCTTCTCCCTGAGTACCGAGTCTTCTCAAAAATGAGAAAGACAATATATATC 179
|||||
61 GCGCCTTCTCCCTGAGTACCGAGTCTTCTCAAAAATGAGAAAGACAATATATATC 120
|||||
180 CATTTCATGATATTCATTAATTAATGAGATATGATGTTTACATGATGATGATGAT 239
|||||
121 CATTTCATGATATTCATTAATTAATGAGATATGATGTTTACATGATGATGATGAT 180
|||||
240 CAGCTGCTATATGCAAAAATGAGATGCTTACAAAGACCCCTTAATTAATTAAC 299
|||||
181 CAGCTGCTATATGCAAAAATGAGATGCTTACAAAGACCCCTTAATTAATTAAC 240
|||||
300 AAGATGGAAGAAAGAAAGAACTCGATGCTGGAATTTGTTCCCTATTAAGATATA 359
|||||
241 AAGATGGAAGAAAGAAAGAACTCGATGCTGGAATTTGTTCCCTATTAAGATATA 300
|||||
360 TCTGAACATATGCTGCTATCCCTGAGACTTGGAAGACCAAGGACCAATGATTAACATA 419

Db 301 TCTGAGCATATGTTGGCATCCCTGACATCTGGAGAACCCAGGCCCAATGATTAACATTA 360
QY 420 CTGGCTGTTGTGTGACAAATGACCAATGATGTTGTGAATTTGGAAGCAAGATATGTG 479
Db 361 CTGGCTGTTGTGTGACAAATGACCAATGATGTTGTGAATTTGGAAGCAAGATATGTG 420
QY 480 CAAGAGTAAATTAATGCGTGAAGTTCTAGCATATTTGGTATGATTTAGCAAGAGGG 539
Db 421 CAAGAGTAAATTAATTTGGCGTGAAGTTCTAGCATATTTGGTATGATTTAGCAAGAGGG 480
QY 540 AAACCGACTGGAAGTCAATTCATTAATGATGATGATCTGATGACCAATTAATG 599
Db 481 AAACCGACTGGAAGTCAATTCATTAATGATGATGATCTGATGACCAATTAATG 540
QY 600 ATATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 626
Db 541 ATATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 567

RESULT 23
AM840924/c 702 bp mRNA EST 18-MAY-2000
LOCUS AM840924 R01-CN0008-290300-014-d09 CN0008 Homo sapiens cDNA, mRNA sequence.
DEFINITION AM840924
ACCESSION AM840924
VERSION AM840924.1 GI:7934907
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 702)

REFERENCE
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=4t2-R01-CN0008-290300-014-d09&t3=2000-03-29&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 649.

FEATURES

SOURCE

1. 702
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CN0008"
/dev_stage="Adult"
/note="Organ: colon, normal; Vector: puc18; Site: 1: Sma1; Site: 2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 203 a 147 c 121 g 231 t
ORIGIN

Query Match 44.5%; Score 567; DB 10; Length 702;
Best Local Similarity 100.0%; Pred. No. 9e-282;
Matches 567; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 541 AACCGACTGGAAGTCAATTCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
Db 595 AACCGACTGGAAGTCAATTCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 536
QY 601 TATCAATG 660
Db 535 TATCAATG 476
QY 661 AAGGATTAAGTTCCTGATGAGAAACAGAAATAGTTCCTGATTAAGCAAGATTA 720
Db 475 AAGGATTAAGTTCCTGATGAGAAACAGAAATAGTTCCTGATTAAGCAAGATTA 416
QY 721 AGATAAGCATTTGGCATGATTAATTAAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
Db 415 AGATAAGCATTTGGCATGATTAATTAAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 356
QY 781 GACTAAGAAAAGCAATGGAAGAAATG 840
Db 355 GACTAAGAAAAGCAATGGAAGAAATG 296
QY 841 CTTCAAGTGTATTCCTGATG 900
Db 295 CTTCAAGTGTATTCCTGATG 236
QY 901 ATCTGCTGCACAGTACCAACAGAGCTGATG 960
Db 235 ATCTGCTGCACAGTACCAACAGAGCTGATG 176
QY 961 AGATTTCCTGGAATACCAAGCTGATG 1020
Db 175 AGATTTCCTGGAATACCAAGCTGATG 116
QY 1021 TAAAGTGTACTTTTCAACCTTAATTTGTAAGTCTATCTACTGATTAAGTATTC 1080
Db 115 TAAAGTGTACTTTTCAACCTTAATTTGTAAGTCTATCTACTGATTAAGTATTC 56
QY 1081 TCGCTGACTAATCCATATATCTCAGA 1107
Db 55 TCGCTGACTAATCCATATATCTCAGA 29

RESULT 24
BGI179769 1076 bp mRNA EST 06-FEB-2001
LOCUS BGI179769
DEFINITION 602328856r1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4430437 5',
mRNA sequence.
ACCESSION BGI179769
VERSION BGI179769.1 GI:12686472
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1076)

REFERENCE NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/OTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Inyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: LRAM10183 row: o column: 14
High quality sequence stop: 667.
Location/Qualifiers

FEATURES

Db 81 TTGTCATAGATGTCAGAGGAGACATATTCGATGATATATGATACCATATGTTAGSAA 22

QY 1259 ATAAATATTTTCTG 1275

Db 21 ATAAATATTTTCTG 5

RESULT 26

AM469221/c EST 24-FEB-2000

LOCUS hc79609.x1 NCI-CGAP.Gas4 Homo sapiens cDNA clone IMAGE:2896932 3'

DEFINITION similar to SW:119YR_BOVIN P37980 INORGANIC PYROPHOSPHATASE ;, mRNA sequence.

ACCESSION AM469221

VERSION AM469221.1

KEYWORDS GI:7039327

SOURCE EST.

ORGANISM human.

Human sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 589)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

Unpublished (1997)

CONTACT: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emert-Buck, M.D., Ph.D.

CNA Library Preparation: Life Technologies, Inc.

CNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.lnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco

High quality sequence stop: 415.

Location/Qualifiers

1. 589

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2896932"

/clone_lib="NCI-CGAP.Gas4"

/tissue_type="poorly differentiated adenocarcinoma with signet ring cell features"

/lab_host="DH10B"

/note="Organ: Stomach; Vector: pCMV-Sport6; Site:1: Salt; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.69 kb. Life Technologies catalog #: 11549-011"

BASE COUNT 179 a 111 c 99 g 199 t 1 others

ORIGIN

Query Match 43.5%; Score 555; DB 10; Length 589;

Best Local Similarity 100.0%; Pred. No. 1.4e-275;

Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 721 AGATAAGACATTGTCATGATATTTTAAAGACATCATGACCATTTGGAACCATTAAT 780

Db 560 AGATAAGACATTGTCATGATATTTTAAAGACATCATGACCATTTGGAACCATTAAT 501

QY 781 GACTAAGAAAGCAATGAAAGCAATGTCATGATATCAACTTTGCTGAGAGCCC 840

Db 500 GACTAAGAAAGCAATGAAAGCAATGTCATGATATCAACTTTGCTGAGAGCCC 441

QY 841 CTTCAGTGTGATCTGATGCTCCAGAGCCATTTGGATGCTTTACCAACCCCTGTGA 900

Db 440 CTTCAGTGTGATCTGATGCTCCAGAGCCATTTGGATGCTTTACCAACCCCTGTGA 381

QY 901 ATCTGCTGTGACAGTACCAACAGCGTATGCTGTTCCATCACCAGAAACCAATG 960

Db 380 ATCTGCTGTGACAGTACCAACAGCGTATGCTGTTCCATCACCAGAAACCAATG 321

QY 961 AGATTCTCTGGAATACAGCTGATATTGCTACATGCTGTTACATCTGATGATTAAG 1020

Db 320 AGATTCTCTGGAATACAGCTGATATTGCTACATGCTGTTACATCTGATGATTAAG 261

QY 1021 TAAAGTAGTAGCTTTTCAAGCTTTAAATTTGTAGACATCTAATCAAGTAATTC 1080

Db 260 TAAAGTAGTAGCTTTTCAAGCTTTAAATTTGTAGACATCTAATCAAGTAATTC 201

QY 1081 TCGTGTAGTAATCCATATATACAGATGATGATTCATCAATCAAGCATTTTCATATCTCA 1140

Db 200 TCGTGTAGTAATCCATATATACAGATGATGATTCATCAATCAAGCATTTTCATATCTCA 141

QY 1141 ACTAAGATTAACCTTTTACACATGCTTAAATATCAAGACAGTGTCAATTTGGAAGTCACTT 1200

Db 140 ACTAAGATTAACCTTTTACACATGCTTAAATATCAAGACAGTGTCAATTTGGAAGTCACTT 81

QY 1201 GTGATAGATGTGCAAGGCGACATATGATGATATGTTACATATGTTAGAAAT 1260

Db 80 GTGATAGATGTGCAAGGCGACATATGATGATATGTTACATATGTTAGAAAT 21

QY 1261 AAAATTAATTTGCTG 1275

Db 20 AAAATTAATTTGCTG 6

RESULT 27

BE867111 EST 20-OCT-2000

LOCUS BE867111

DEFINITION 60144295.F1 NIH-MGC_65 Homo sapiens cDNA clone IMAGE:3846924 5', mRNA sequence.

ACCESSION BE867111

VERSION BE867111.1

KEYWORDS GI:10315887

SOURCE EST.

ORGANISM human.

Human sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 592)

NIH-MGC http://imgc.ncbi.nlm.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

CNA Library Preparation: Life Technologies, Inc.

CNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: L1AM9560 row: f column: 13

High quality sequence stop: 592.

Location/Qualifiers

1. 592

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3846924"

/clone_lib="NIH-MGC_65"

/tissue_type="adenocarcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: Colon; Vector: pCMV-Sport6; Site:1: NotI; Site:2: Salt; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."

BASE COUNT 174 a 117 c 154 g 147 t

ORIGIN

Query Match 43.3%; Score 552; DB 11; Length 592;

Best Local Similarity 100.0%; Pred. No. 5e-274;

Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 79 CTCGGCAGCTATGAGCGGCTTCAGCAGCAGAGCGCGCCGCTTCCTCCAGAGTA 138
    |||||||
Db 41 CTCGGCAGCTATGAGCGGCTTCAGCAGCAGAGCGCGCCGCTTCCTCCAGAGTA 100
OY 139 CCGAGCTCTTCTCAAAAAATGAGAAAGCAATATATATCTCCATTTTCATGATATTCAT 198
    |||||||
Db 101 CCGAGCTCTTCTCAAAAAATGAGAAAGCAATATATATCTCCATTTTCATGATATTCAT 160
OY 199 TTATGACAGATAGAGATGTGTTTCACATGTGTGTTGAGTACCGCGCTGCTTAATGCATA 258
    |||||||
Db 161 TTATGACAGATAGAGATGTGTTTCACATGTGTGTTGAGTACCGCGCTGCTTAATGCATA 220
OY 259 AATGAGATGCTCTCAAAAGACCCCTTTAAACCTATTAAACAGATGTGAAAAAAGAGAA 318
    |||||||
Db 221 AATGAGATGCTCTCAAAAGACCCCTTTAAACCTATTAAACAGATGTGAAAAAAGAGAA 280
OY 319 ACTTCGCTATGTGCGAATTTTGTCCGATATTAAGATATATCTGGAACATATGTGCCAT 378
    |||||||
Db 281 ACTTCGCTATGTGCGAATTTTGTCCGATATTAAGATATATCTGGAACATATGTGCCAT 340
OY 379 CCCTCAGACTTGGAAGACCCAGGCGCAATGATTAACATCTGCGCTGTGTGCGACAA 438
    |||||||
Db 341 CCCTCAGACTTGGAAGACCCAGGCGCAATGATTAACATCTGCGCTGTGTGCGACAA 400
OY 439 TGACCCAAATGATGTGTGTAATTTGAAAGCAAGTATGTCAAGAGGTGAATTAATGG 498
    |||||||
Db 401 TGACCCAAATGATGTGTGTAATTTGAAAGCAAGTATGTCAAGAGGTGAATTAATGG 460
OY 499 CGTGAAGTTCTAGGCAATATTTGCTATGATTGACGAAGGGAACCGACGTAAGATCAT 558
    |||||||
Db 461 CGTGAAGTTCTAGGCAATATTTGCTATGATTGACGAAGGGAACCGACGTAAGATCAT 520
OY 559 TGGCATTAAATGATGATGCTGATGCGACCAATTAATATATCATATGATGTAACG 618
    |||||||
Db 521 TGGCATTAAATGATGATGCTGATGCGACCAATTAATATATCATATGATGTAACG 580
OY 619 GCTGAACCTGG 630
    |||||||
Db 581 GCTGAACCTGG 592

RESULT 28
Bg721093 762 bp mRNA EST 08-MAY-2001
LOCUS 602692932E1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4825206 5'
DEFINITION mRNA sequence.
ACCESSION Bg721093
VERSION Bg721093.1 GI:14000280
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 762)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
            Email: rgabbs-remail.nih.gov
            Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNLN at:
            http://image.llnl.gov
            Plate: LRAM10737 row: P column: 07
            High quality sequence stop: 753.
            Location/Qualifiers
            1..762
            /organism="Homo sapiens"
            /db_xref="taxon:9606"

```

```

/clone="IMAGE:4825206"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescript (modified
pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH-MGC Library."
BASE COUNT 230 a 140 c 197 g 194 t 1 others
ORIGIN

Query Match 42.7%; Score 544; DB 11; Length 762;
Best Local Similarity 99.8%; Pred. No. 6.9e-270;
Matches 594; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 25 GTCAATCGCGCGCGCTGTGGGCTGTGCTCTGTGCGACGCGCGCGCAGACGCTCCG 84
    |||||||
Db 6 GTCAATCGCGCGCGCTGTGGGCTGTGCTCTGTGCGACGCGCGCGCAGACGCTCCG 65
OY 85 CACTATGAGCGGCTTCAGACAGCAGAGCGCGCGCGCTTCCTCCCTGAGTACCGAGT 144
    |||||||
Db 66 CACTATGAGCGGCTTCAGACAGCAGAGCGCGCGCGCTTCCTCCCTGAGTACCGAGT 125
OY 145 CTTCCTCAAAAATGAGAAAGCAATATATATCTCCATTTTCATGATATTCATTTATC 204
    |||||||
Db 126 CTTCCTCAAAAATGAGAAAGCAATATATATCTCCATTTTCATGATATTCATTTATC 185
OY 205 AGATAGAGATGTGTTTCACATGTGTGTTGAGTACACCGCTGTCTTAATGCAAAATGGA 264
    |||||||
Db 186 AGATAGAGATGTGTTTCACATGTGTGTTGAGTACACCGCTGTCTTAATGCAAAATGGA 245
OY 265 GATTGCTCAAAAGGACCCCTTTAAACCTATTAACAGATGTGAAAAAAGAAACTGG 324
    |||||||
Db 246 GATTGCTCAAAAGGACCCCTTTAAACCTATTAACAGATGTGAAAAAAGAAACTGG 305
OY 325 CTATGTGGGAATTTGTTCCGCTATTAAGATATATCTGGAACATATGTCCTCATC 384
    |||||||
Db 306 CTATGTGGGAATTTGTTCCGCTATTAAGATATATCTGGAACATATGTCCTCATC 365
OY 385 GACTTGGGAAGACCCAGGCGCAATGATTAACATATCTGCTGTGTGCGACAAATGACC 444
    |||||||
Db 366 GACTTGGGAAGACCCAGGCGCAATGATTAACATATCTGCTGTGTGCGACAAATGACC 425
OY 445 AATTGATGTGTGAAATTTGGAAGCAAGTATGTGCAAGAGGTGAATATTTGCGCTGAA 504
    |||||||
Db 426 AATTGATGTGTGAAATTTGGAAGCAAGTATGTGCAAGAGGTGAATATTTGCGCTGAA 485
OY 505 AGTTCTAGGCAATTTGGCTATGATGACGAGGGAACCGACGTGAAGTCTTCCTCAT 564
    |||||||
Db 486 AGTTCTAGGCAATTTGGCTATGATGACGAGGGAACCGACGTGAAGTCTTCCTCAT 545
OY 565 TAATGCGATGATCCGATGACGACCAATTAATATGATATCATGATGCTCAACCG 619
    |||||||
Db 546 TAATGCGATGATCCGATGACGACCAATTAATATGATATCATGATGCTCAACCG 600

RESULT 29
AA694204/c 639 bp mRNA EST 16-DEC-1997
LOCUS z142c12.s1 Soares_fetal_liver.spleen_INFUS.S1 Homo sapiens cDNA
DEFINITION clone IMAGE:433462 3' similar to SW:IPYR_BOVIN P37980 INORGANIC
ACCESSION AA694204
VERSION AA694204.1 GI:2695142
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```


Db	563	CCATGATATTTAAAGACACCTCAGACCATTTGGAAAGCATTACTACTAAGAAAGCA	504
OY	795	ATGAAAGAAATCAGTTCGATGATACATACACTTTGTCTGAGAGCCCTTCAGTGTGATC	854
Db	503	ATGAAAGAAATCAGTTCGATGATACATACACTTTGTCTGAGAGCCCTTCAGTGTGATC	444
OY	855	CTGATGCTGCACAGACCATTTGTGATGCTTTACACACACCCGTGAATCTGCTGCACAG	914
Db	443	CTGATGCTGCACAGACCATTTGTGATGCTTTACACACACCCGTGAATCTGCTGCACAG	384
OY	915	TACCAACAGACGTGATTAAGTGGTTCATCACCAGAAAACTAATGAGATTTCTCTGAA	974
Db	383	TACCAACAGACGTGATTAAGTGGTTCATCACCAGAAAACTAATGAGATTTCTCTGAA	324
OY	975	TACACGTGATATTTGCTACTACATCGTGTTCATCTGGATGTTTGAAGTAAAGTAGTACT	1034
Db	323	TACACGTGATATTTGCTACTACATCGTGTTCATCTGGATGTTTGAAGTAAAGTAGTACT	264
OY	1035	TTTCAAGCTTTAAATTTGTAGAACCTCACTTAACCTAAATTTCTGCTGTGACTATC	1094
Db	263	TTTCAAGCTTTAAATTTGTAGAACCTCACTTAACCTAAATTTCTGCTGTGACTATC	204
OY	1095	CAATATCTCAGATGTTATCCATCCATTAAGCAATTTTCTATCTCACTAAGATTAATCTT	1154
Db	203	CAATATCTCAGATGTTATCCATCCATTAAGCAATTTTCTATCTCACTAAGATTAATCTT	144
OY	1155	TAGCACATGCTTAATATCAAAACAGTTTGCATTTTGGAGTAGTCTTGATATAGATGTGC	1214
Db	143	TAGCACATGCTTAATATCAAAACAGTTTGCATTTTGGAGTAGTCTTGATATAGATGTGC	84
OY	1215	AAGGGAGCACATATTTGGATGTAATATGTTACCATATGTTAGGAAATTAATAATTTTGGCT	1274
Db	83	AAGGGAGCACATATTTGGATGTAATATGTTACCATATGTTAGGAAATTAATAATTTTGGCT	24
OY	1275	G 1275	
Db	23	G 23	

RESULT 31
 BG704401
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

BG704401 750 bp mRNA EST 07-MAY-2001
 602687468F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4820008 5',
 mRNA sequence.
 BG704401
 BG704401.1 GI:13977706
 EST.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; catarrhini; Hominiidae; Homo.
 1 (bases 1 to 750)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabds-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitlak
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.jnl.gov>
 plate: LLM10724 row: g column: 17
 High quality sequence stop: 744.

FEATURES	location/Qualifiers
source	1..750
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:4820008"
	/clone_lib="NIH_MGC_95"

/tissue_type="hippocampus"
/lab_host="DH10B"
pbhscriptr KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtagag
Oligo-dn primed using primer 5'-TTTTTGGTTTCTTTTTCVN-3',
size-selected for average insert size 2.5 kb and
normalized to RQF 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHRRI, National
Institutes of Health). Note: this is a NIH-MGC Library.
BASE COUNT 224 a 139 c 195 g 191 t 1 others
ORIGIN

Query Match	Similarity	Score	DB	Length
Best Local	539	100.0%	2.7e-267	
Matches	539	Conservative	0	Mismatches
			0	Indels
			0	Gaps
				0
42.3%				
Score	539		DB	11
Pred.	No.	2.7e-267		
DB	11			
Length	750			
Query	Match	Similarity	Score	DB
1	23	100.0%	100.0%	11
2	8	100.0%	100.0%	11
3	83	100.0%	100.0%	11
4	68	100.0%	100.0%	11
5	143	100.0%	100.0%	11
6	128	100.0%	100.0%	11
7	203	100.0%	100.0%	11
8	188	100.0%	100.0%	11
9	263	100.0%	100.0%	11
10	248	100.0%	100.0%	11
11	323	100.0%	100.0%	11
12	308	100.0%	100.0%	11
13	383	100.0%	100.0%	11
14	368	100.0%	100.0%	11
15	443	100.0%	100.0%	11
16	428	100.0%	100.0%	11
17	503	100.0%	100.0%	11
18	488	100.0%	100.0%	11

RESULT	32
LOCUS	BG714352
DEFINITION	BG714352 593 bp mRNA EST 08-MAY-2001 602665986f1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:479290 5' , mRNA sequence.
ACCESSION	BG714352
VERSION	BG714352.1 GI:13993283
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 593) NIH-MGC http://mgc.ncl.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) unpublished (1999)
AUTHORS	Contact: Robert Strausberg, Ph.D.
TITLE	
JOURNAL	
COMMENT	

Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (NLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/NLN at:
<http://image.llnl.gov>
Plate: LLM10671 row: 1 column: 15
High quality sequence stop: 591.
Location/Qualifiers

BASE COUNT	167 a	119 c	158 g	149 t
ORIGIN				

Query Match	42.0%	Score 535;	DB 11;	Length 593;
Best Local Similarity	99.8%;	Pred. No. 3e-265;		
Matches 585;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

QY	18	TTCTCTGTGTAGTGGGGCGCGGCTGCGGGGCTGTGGCTGTGGTACAGAGCGGGCGGGACAG	77
Db	8	TTCTCTGTGTAGTGGGGCGCGGCTGCGGGGCTGTGGCTGTGTGGACAGCGCGCGGACAG	67
QY	78	ACTCCGGCAGTATGAGCGGGCTTCAGCACCGAGGAGCGCGCGGCGCTTCCTCTGGAGT	137
Db	68	ACTCCGGCAGTATGAGCGGGCTTCAGCACCGAGGAGCGCGCGGCGCTTCCTCTGGAGT	127
QY	138	ACCGAGTCTTCTCTCAAAAATGAGAAAGGACAATATATCTCCATTTCATGATATTCGAA	197
Db	128	ACCGAGTCTTCTCTCAAAAATGAGAAAGGACAATATATCTCCATTTCATGATATTCGAA	187
QY	198	TTTATGCGAATAGAGATGTGTTTCACATGCTGATGGAAGTACACAGCTGGTCTAATGCAA	257
Db	188	TTTATGCGAATAGAGATGTGTTTCACATGCTGATGGAAGTACACAGCTGGTCTAATGCAA	247
QY	258	AAATGAGATTTGCTTCAAAAGGACCCCTTTAAACCCCTATTAAACAAGATGTGAAAAAGGAA	317
Db	248	AAATGAGATTTGCTTCAAAAGGACCCCTTTAAACCCCTATTAAACAAGATGTGAAAAAGGAA	307
QY	318	AACTTCGGTATGTCGCAATTTGTTCCCGTATAAAGATATATCTGGAACATAATGTGCGCA	377
Db	308	AACTTCGGTATGTCGCAATTTGTTCCCGTATAAAGATATATCTGGAACATAATGTGCGCA	367
QY	378	TCCCTCACACTTGGGANAAGCCGAGGGACAAATGATATAACATCTGGCTGTGGGTGACA	437
Db	368	TCCCTCACACTTGGGANAAGCCGAGGGACAAATGATATAACATCTGGCTGTGGGTGACA	427
QY	438	ATGACCAATTGATGTGTGTGAAATGTGAAGCAAGCAGGTATGTCCAMAGSGTGAATATATG	497
Db	428	ATGACCAATTGATGTGTGTGAAATGTGAAGCAAGGTATGTCCAMAGSGTGAATATATG	487
QY	498	GGGTAAAGTGTATGACCAATTTGGCTATGATGACGAAGGGGAAACCGACTGGAAAGTCA	557
Db	488	GGGTAAAGTGTATGACCAATTTGGCTATGATGACGAAGGGGAAACCGACTGGAAAGTCA	547
QY	558	TTGCATTAAATGTGATGTATCTGTGATGCAAGCCCAATTAATATATATAT 603	

```

Db      548   TTGCATTATGTGATGATCCTGATGCCAGCCAAATATAATGATAT 593
|||||
RESULT .33
BF033412
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BF033412          914 bp      mRNA
601457932F1 NIH_MGC_66 Homo sapiens CDNA clone IMAGE:3861727 5',
mRNA sequence.
BF033412
BF033412.1 GI:10741124
EST
human.
Homo Sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 914)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL).
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLAM9598 row: o column: 08
High quality sequence stop: 611.
```

BASE COUNT	277 a	172 c	225 g	240 t	
ORIGIN					
Query Match		41.7%	Score 532;	DB 11;	Length 914;
Best Local Similarity		99.8%	Pred. No. 11e-26;		
Matches 582; Conservative		0;	Mismatches 1;	Indels 0;	Gaps 0;

Query Match	41.7%	Score 532	DB 11	Length 914
Best Local Similarity	99.8%	Pred: N.1le-26;		
Matches 582	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY 184	TCATGATATTCGAATTTATGACAGATPAGCATGTGTTCCATCATGTAGTGAAGTACCACG	243		
Db 1	TCATGATATTCGAATTTATGACAGATPAGCATGTGTTCCATCATGTAGTGAAGTACCACG	50		
QY 244	CTGGTCTAATGCAAAAAATGAGATTGCTTCAAAAGACCCCTTTAAACCCATTTAAACAA	303		
Db 61	CTGGTCTAATGCAAAAAATGAGATTGCTTCAAAAGACCCCTTTAAACCCATTTAAACAA	120		
QY 304	TGTGAAAAAAGGAAACCTTCGTATGTTGCGAATTTGTTCCCGTATAAAGATATTTCTG	363		
Db 121	TGTGAAAAAAGGAAACCTTCGTATGTTGCGAATTTGTTCCCGTATAAAGATATTTCTG	180		
QY 364	GAACATATGTGCCATCCCTCAGACTTGGGAAGCCAGGCGACATGATTAACATCTCG	423		
Db 181	GAACATATGTGCCATCCCTCAGACTTGGGAAGCCAGGCGACATGATTAACATCTCG	240		
QY 424	CTGTATGGGACAATGACCCATTTGATGTGTGAAATTTGGAAGCAAGGTATGTGCAAG	483		
Db 241	CTGTATGGGACAATGACCCATTTGATGTGTGAAATTTGGAAGCAAGGTATGTGCAAG	300		
QY 484	AGGTGAATTAATTTGGCGTGAACCTTACGACATATTTGCTATGATTTGACGAAGGGAAC	543		

Db 301 AGTGAAATATATGGCGTGAAGTTCTAGGCATATGGCTATGATTCAGCAAGGGGAAC 360
QY 544 CGACTGGAAGTCAATGGCATTAATGTGATGATCTGATGACAGCAATTAATGATAT 603
Db 361 CGACTGGAAGTCAATGGCATTAATGTGATGATCTGATGACAGCAATTAATGATAT 420
QY 604 CAAATGATGTCAGGAGGCTGGAAGCTGGCTACTTACAGAGCTACTGTGAGCTGTTAGAG 663
Db 421 CAAATGATGTCAGGAGGCTGGAAGCTGGCTACTTACAGAGCTACTGTGAGCTGTTAGAG 480
QY 664 GTATAGAGTTCCTGATGGAAGCAAGAAATGAGTTCGTTAATGAGATTAAGA 723
Db 481 GTATAGAGTTCCTGATGGAAGCAAGAAATGAGTTCGTTAATGAGATTAAGA 540
QY 724 TAAGACTTTGTCATGATATTAATAAGCACTCATGACCAT 766
Db 541 TAAGACTTTGTCATGATATTAATAAGCACTCATGACCAT 583

RESULT 34
AM518147/c 631 bp mRNA EST 03-MAR-2000
LOCUS xx81f10.x1 NCI_CGAP_Lym12 Homo sapiens cDNA clone IMAGE:2850091 3'
DEFINITION similar to SW:IPYR_BOVIN P37980 INORGANIC PYROPHOSPHATASE ;, mRNA
ACCESSION AM518147
VERSION AM518147
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 631)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Life Technologies catalog #: 11547-015
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/lresources.shtml

FEATURES
source
1..631
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2850091"
/clone_lib="NCI_CGAP_Lym12"
/tissue_type="lymphoma, follicular mixed small and large
cell"
/lab_host="DH10B"
/note="Organ: lymph node; Vector: pCMV-SPORT6; Site:1:
Salt: Site:2: NotI; Cloned unidirectionally. Primer:
Oligo dT. Average insert size 1.25 kb. Life Technologies
catalog #: 11547-015"
BASE COUNT 188 a 123 c 102 g 217 t 1 others
ORIGIN
Possible reversed clone: polyT not found
Seq primer: -400P from GIBCO
High quality sequence stop: 409.
Location/Qualifiers
1..631

FEATURES
source
1..570
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1883873"
/clone_lib="NCI_CGAP_Lu5"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pRT3D-Pac (Pharmacia) with a

Query Match 41.4%; Score 528; DB 10; Length 631;
Best Local Similarity 100.0%; Prid. No. 1.3e-261;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 748 TAAAGCAGCTCATGACCTTGAAGCACTTAGTGTAGCTAAGAAAGCATGGAAGGAAT 807
Db 534 TAAAGCAGCTCATGACCTTGAAGCACTTAGTGTAGCTAAGAAAGCATGGAAGGAAT 475

QY 808 CAGTTCATGATATACACTTTTCTGAGAGCCCTTCAAGTGTATCTGATGCTGCCAG 867
Db 474 CAGTTCATGATATACACTTTTCTGAGAGCCCTTCAAGTGTATCTGATGCTGCCAG 415
QY 868 AGCCATTGGATGGTTTACACCAACCCGTGAACTGCTGCACAGTACCAAGAGCT 927
Db 414 AGCCATTGGATGGTTTACACCAACCCGTGAACTGCTGCACAGTACCAAGAGCT 355
QY 928 GGATAGTGGTTCCTACACACAGAAAACATAGATTTCTGTGAAATCAAGTGTAT 987
Db 354 GGATAGTGGTTCCTACACACAGAAAACATAGATTTCTGTGAAATCAAGTGTAT 295
QY 988 TCTACATGCTGTTTATCTGATGATTAAGAAAGTAAAGTACTAGCTTTTCAAGCTTTA 1047
Db 294 TCTACATGCTGTTTATCTGATGATTAAGAAAGTAAAGTACTAGCTTTTCAAGCTTTA 235
QY 1048 AATTGTAGACTCATCTACCTAAGTAATTTCTGCTGACTAATCCATATACCTAGA 1107
Db 234 AATTGTAGACTCATCTAAGTAATTTCTGCTGACTAATCCATATACCTAGA 175
QY 1108 ATGTTATCATCTAAGCAATTTTTCATATCTGACCTAAGTACTTTTACACATGCTTA 1167
Db 174 ATGTTATCATCTAAGCAATTTTTCATATCTGACCTAAGTACTTTTACACATGCTTA 115
QY 1168 AATATCAAGAGCTGTCTCATTTTGAAGTCACTTGTGAATGATGTGCAAGGAGGACAT 1227
Db 114 AATATCAAGAGCTGTCTCATTTTGAAGTCACTTGTGAATGATGTGCAAGGAGGACAT 55
QY 1228 ATTGATGTATATGTTACCATATGTTAGAAATTAATTTTGTCTG 1275
Db 54 ATTGATGTATATGTTACCATATGTTAGAAATTAATTTTGTCTG 7

RESULT 35
AI215898/c 570 bp mRNA EST 21-DEC-1998
LOCUS qm35h09.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1883873 3'
DEFINITION similar to SW:IPYR_BOVIN P37980 INORGANIC PYROPHOSPHATASE ;, mRNA
ACCESSION AI215898
VERSION AI215898
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 570)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1001 Std Error: 0.00
Seq primer: -400P from GIBCO
High quality sequence stop: 353.
Location/Qualifiers
1..570

FEATURES
source
1..570
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1883873"
/clone_lib="NCI_CGAP_Lu5"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pRT3D-Pac (Pharmacia) with a

Query Match 41.4%; Score 528; DB 10; Length 631;
Best Local Similarity 100.0%; Prid. No. 1.3e-261;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 748 TAAAGCAGCTCATGACCTTGAAGCACTTAGTGTAGCTAAGAAAGCATGGAAGGAAT 807
Db 534 TAAAGCAGCTCATGACCTTGAAGCACTTAGTGTAGCTAAGAAAGCATGGAAGGAAT 475

modified polylinker. 1st strand cDNA was prepared from neuroendocrine lung carcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 172 a 108 c 98 g 192 t
ORIGIN

Query Match 41.0%; Score 523; DB 10; Length 570;
Best Local Similarity 100.0%; Pred. No. 4,8e-259;
Matches 523; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

753 GCACATGACCATGGAAGCATTAGTACTAAGAAAAGCATGGAAGATCACTT 812
|||||
524 GCACATGACCATGGAAGCATTAGTACTAAGAAAAGCATGGAAGATCACTT 465
|||||
813 GCATGATACATCTTGTCTGAGACCCCTTCAAGTGTGATGCTGCAGAGCCA 872
|||||
464 GCATGATACATCTTGTCTGAGACCCCTTCAAGTGTGATGCTGCAGAGCCA 405
|||||
873 TGTGTGATGCTTACACACCCCTGGAATCTGCTGCACAGTACCAAGAGAGTGA 932
|||||
404 TGTGTGATGCTTACACACCCCTGGAATCTGCTGCACAGTACCAAGAGAGTGA 345
|||||
933 AGTGTGATGCTTACACACCCCTGGAATCTGCTGCACAGTACCAAGAGAGTGA 992
|||||
344 AGTGTGATGCTTACACACCCCTGGAATCTGCTGCACAGTACCAAGAGAGTGA 285
|||||
993 CATTGCTGATGCTTACACACCCCTGGAATCTGCTGCACAGTACCAAGAGAGTGA 1052
|||||
284 CATTGCTGATGCTTACACACCCCTGGAATCTGCTGCACAGTACCAAGAGAGTGA 225
|||||
1053 GTAGAACTATCTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGT 1112
|||||
224 GTAGAACTATCTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGT 165
|||||
1113 ATCCATCTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGT 1172
|||||
164 ATCCATCTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGT 105
|||||
1173 CAAGACAGTGTCTATTTGGAAGTCACTGTGAATAGATGCAAGGGAGACATATTTG 1232
|||||
104 CAAGACAGTGTCTATTTGGAAGTCACTGTGAATAGATGCAAGGGAGACATATTTG 45
|||||
1233 ATGTATATGTTACCATATGTTAGAAATTAATTTTGTCTG 1275
|||||
44 ATGTATATGTTACCATATGTTAGAAATTAATTTTGTCTG 2

RESULT 36
AV706813 690 bp mRNA EST 09-OCT-2000
LOCUS AV706813 ADB Homo sapiens cDNA clone ADBCOF06 5', mRNA sequence.
DEFINITION AV706813
ACCESSION AV706813
VERSION AV706813.1 GI:10724091
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 690)
AUTHORS Peng, Y., Song, H., Huang, O., Huang, C., Gu, Y., Yang, Y., Gao, G., Xiao, H., Xu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R., Chen, J., Chen, Z., and Han, Z.
TITLE Homo sapiens cDNA ADB clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China.
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source location/Qualifiers

1. 690
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="ADBCOF06"
/clone_lib="ADB"
/tissue_type="Adrenal gland"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: Bluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 224 a 121 c 153 g 190 t 2 others
ORIGIN

Query Match 40.5%; Score 517; DB 10; Length 690;
Best Local Similarity 100.0%; Pred. No. 6,2e-256;
Matches 517; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

481 AAGAGTGAATTAATTTGGCGTGAAGTCTAGGATATTTGGCTATGAGTGAAGGGA 540
|||||
1 AAGAGTGAATTAATTTGGCGTGAAGTCTAGGATATTTGGCTATGAGTGAAGGGA 60
|||||
541 AAGAGTGAATTAATTTGGCGTGAAGTCTAGGATATTTGGCTATGAGTGAAGGGA 600
|||||
61 AAGAGTGAATTAATTTGGCGTGAAGTCTAGGATATTTGGCTATGAGTGAAGGGA 120
|||||
601 TATCAATGATGTCAGGCGTGAAGTCTAGGATATTTGGCTATGAGTGAAGGGA 660
|||||
121 TATCAATGATGTCAGGCGTGAAGTCTAGGATATTTGGCTATGAGTGAAGGGA 180
|||||
661 AAGATTAAGTTCCTGATGTCAGGCGTGAAGTCTAGGATATTTGGCTATGAGTGAAGGGA 720
|||||
181 AAGATTAAGTTCCTGATGTCAGGCGTGAAGTCTAGGATATTTGGCTATGAGTGAAGGGA 240
|||||
721 AAGATTAAGTTCCTGATGTCAGGCGTGAAGTCTAGGATATTTGGCTATGAGTGAAGGGA 780
|||||
241 AAGATTAAGTTCCTGATGTCAGGCGTGAAGTCTAGGATATTTGGCTATGAGTGAAGGGA 300
|||||
781 GACTAAGAAAGCATGTCAGGCGTGAAGTCTAGGATATTTGGCTATGAGTGAAGGGA 840
|||||
301 GACTAAGAAAGCATGTCAGGCGTGAAGTCTAGGATATTTGGCTATGAGTGAAGGGA 360
|||||
841 CTTCAAGTGTGATCCTGATGTCAGGCGTGAAGTCTAGGATATTTGGCTATGAGTGAAGGGA 900
|||||
361 CTTCAAGTGTGATCCTGATGTCAGGCGTGAAGTCTAGGATATTTGGCTATGAGTGAAGGGA 420
|||||
901 ATCTGCTGTCAGATGTCAGGCGTGAAGTCTAGGATATTTGGCTATGAGTGAAGGGA 960
|||||
421 ATCTGCTGTCAGATGTCAGGCGTGAAGTCTAGGATATTTGGCTATGAGTGAAGGGA 480
|||||
961 AGATTTCTCTGGAATTCAGGCGTGAAGTCTAGGATATTTGGCTATGAGTGAAGGGA 997
|||||
481 AGATTTCTCTGGAATTCAGGCGTGAAGTCTAGGATATTTGGCTATGAGTGAAGGGA 517
|||||

RESULT 37
BG036446 775 bp mRNA EST 24-JAN-2001
LOCUS BG036446 602326583F1 NIH.MGC.91 Homo sapiens cDNA clone IMAGE:4427925 5',
DEFINITION BG036446
ACCESSION BG036446
VERSION BG036446.1 GI:12431619
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

http://image.llnl.gov
Plate: LLAM10305 row: 1 column: 16
High quality sequence stop: 667.

FEATURES

source

Location/Qualifiers

1. 940.

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4477215"

/clone_lib="NIH_MGC_91"

/tissue_type="adeno carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: prostate; Vector: PCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dt primed."

Average insert size 1.4 kb. Library enriched for full-length clones and constructed by Life Technologies.

Note: this is a NIH_MGC library."

BASE COUNT

275 a 183 c 232 g 250 t

ORIGIN

Query Match 39.5%; Score: 504; DB 11; Length 940;

Best Local Similarity 100.0%; Pred. No. 3.3e-249; Mismatches 0; Indels 0; Gaps 0;

Matches 504; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

402 GGCACATGATTAACATGCTGCTGTGTCGACATGACCCCAATGATGTGTGAAA 461

40 GGCACATGATTAACATGCTGCTGTGTCGACATGACCCCAATGATGTGTGAAA 99

462 TTGGAGCAAGGTATGTGCAAGAGGTGAATATTTGGCGTGAAGCTTCTAGCATTTGG 521

100 TTGGAGCAAGGTATGTGCAAGAGGTGAATATTTGGCGTGAAGCTTCTAGCATTTGG 159

522 CTATGATGACGAAGGGAACGACATGGAAGCATTTGCCATTAATGTGATGATCTG 581

160 CTATGATGACGAAGGGAACGACATGGAAGCATTTGCCATTAATGTGATGATCTG 219

582 ATGACGCAATTAATGATATCATGATGTCACACGCTGAACCTGCTACTTAAAG 641

220 ATGACGCAATTAATGATATCATGATGTCACACGCTGAACCTGCTACTTAAAG 279

642 CTACTGTGAGCTGTTAGAAAGGTATAGTCTCTGATGCAAAACCGAAATAGTTTG 701

280 CTACTGTGAGCTGTTAGAAAGGTATAGTCTCTGATGCAAAACCGAAATAGTTTG 339

702 CGTTTAATGAGCAATTAAGATTAAGACATTTGCCATTAATTAATAAAGCATCTG 761

340 CGTTTAATGAGCAATTAAGATTAAGACATTTGCCATTAATTAATAAAGCATCTG 399

762 ACCATTGGAAGCAATTAAGATTAAGCAATGGAAGCAATGATGATGATGATA 821

400 ACCATTGGAAGCAATTAAGATTAAGCAATGGAAGCAATGATGATGATGATA 459

822 CAACCTTGTGTGAGAGCCCTTCAAGTGTGATCTGATGCTGCGAGAGCCATTTGATG 881

460 CAACCTTGTGTGAGAGCCCTTCAAGTGTGATCTGATGCTGCGAGAGCCATTTGATG 519

882 CTTTACCACCACTGATCTG 905

520 CTTTACCACCACTGATCTG 543

RESULT 41

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BE887538 503 bp mRNA EST 20-OCT-2000

601508191P1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3909526 5'

mRNA sequence.

BE887538.1 GI:10342930

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 503)

AUTHORS NIH-MGC http://imgc.ncl.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaabs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library: Preparation: Life Technologies, Inc.

CDNA Library: Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM9723 row: f column: 23

High quality sequence stop: 503.

FEATURES

source

Location/Qualifiers

1. 503

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3909526"

/clone_lib="NIH_MGC_71"

/tissue_type="leiomysarcoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: PCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: oligo dt."

Average insert size 2.1 kb.

BASE COUNT

154 a 97 c 126 g 126 t

ORIGIN

Query Match 39.5%; Score: 503; DB 11; Length 503;

Best Local Similarity 100.0%; Pred. No. 1e-248; Mismatches 0; Indels 0; Gaps 0;

Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

69 GCGGCGAGACTCCGCGCTATGACGGCTTCAGCAGGAGAGCGCGCCCTTCT 128

1 GCGGCGAGACTCCGCGCTATGACGGCTTCAGCAGGAGAGCGCGCCCTTCT 60

129 CCGTGAGTACGAGTCTTCCCAAAATGAGAAAGCAATATATTCATTTGATG 188

61 CCGTGAGTACGAGTCTTCCCAAAATGAGAAAGCAATATATTCATTTGATG 120

189 AATATCCAAATTTATGACATTAAGATGTTTCAATGATGATGATGATGATGATG 248

121 AATATCCAAATTTATGACATTAAGATGTTTCAATGATGATGATGATGATGATG 180

249 CTAATGCAAAATGAGATGTTGCTACAAAGACCTTTAAACCTTATTAACAGATGCA 308

181 CTAATGCAAAATGAGATGTTGCTACAAAGACCTTTAAACCTTATTAACAGATGCA 240

309 AAAAAGCAAACTTCGATGTTGCGAATTTGTCCTATTAAGCATATATCTGAACT 368

241 AAAAAGCAAACTTCGATGTTGCGAATTTGTCCTATTAAGCATATATCTGAACT 300

369 ATGGTCCATCCCTCAGACTTGGGAAGACCGAGGACATGATTAACATGCTGCTGT 428

301 ATGGTCCATCCCTCAGACTTGGGAAGACCGAGGACATGATTAACATGCTGCTGT 360

429 GTGTGACAAATGACCAATGATGTTGTAATTTGAAAGCAAGATGATGCAAGAGTG 488

361 GTGTGACAAATGACCAATGATGTTGTAATTTGAAAGCAAGATGATGCAAGAGTG 420

489 AATATATGGCGTGAATTTCTAGCAATTTGGCTATATGACGAAGGGGAAACCGACT 548

421 AATATATGGCGTGAATTTCTAGCAATTTGGCTATATGACGAAGGGGAAACCGACT 480

549 GGAAGCATGTCATTAATG 571

481 GGAAGCATGTCATTAATG 503

RESULT 42

BE042554/c

LOCUS	BE042554	585 bp	mtDNA	EST	08-JUN-2000
DEFINITION	ho25e08.x1 NCI-CGAP.Co14 Homo sapiens CDNA clone IMAGE:3038438 3' similar to SW:IPYR_BOVIN P37980 INORGANIC PYROPHOSPHATASE ;, mtDNA sequence.				
ACCESSION	BE042554				
VERSION	BE042554.1				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
JOURNAL	Unpublished (1997)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov Seq primer: -400P from Gibbo High quality sequence stop: 415. Location/Qualifiers				
FEATURES	Source				
	1..585				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:3038438"				
	/clone_lib="NCI CGAP.Co14"				
	/tissue_type="moderately-differentiated adenocarcinoma"				
	/lab_host="DH10B"				
	/note="Organ: colon; Vector: pCMV-SPORT6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert. size 1.7 kb. Life Technologies catalog #: 11531-019"				
BASE COUNT	175 a 106 c 99 g 204 t 1 others				
ORIGIN					
Query Match	39.5%; Score 503; DB 10; Length 585;				
Best Local Similarity	100.0%; Pred. No. 1e-248;				
Matches 503; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
773	GCATTAGTACATCAAGAAACGAATGAAAGAAATCAGTGTGATGAAATCACTTTGCT	832			
515	GCATTAGTACATCAAGAAACGAATGAAAGAAATCAGTGTGATGAAATCACTTTGCT	456			
833	GAGAGCCCTTCAAGTGTGATCCTGATGCTGCCAGAGCCATTGTGGATGCTTTACACCA	892			
455	GAGAGCCCTTCAAGTGTGATCCTGATGCTGCCAGAGCCATTGTGGATGCTTTACACCA	396			
893	CCCTGGAATTCGCCCGCAGCAGTACCAACAGACGTGATTAAGTGGTTCATCACACAGAA	952			
395	CCCTGGAATTCGCCCGCAGCAGTACCAACAGACGTGATTAAGTGGTTCATCACACAGAA	336			
953	AACATAAGATTTCTCTGGAATACAAAGCTGATATTGCTACATCGTTCATCTGGATGT	1012			
335	AACATAAGATTTCTCTGGAATACAAAGCTGATATTGCTACATCGTTCATCTGGATGT	276			
1013	ATTAGAAGTAAAGTAGTAGCTTTTCAAGCTTTAAATTTGTAGAACATCTAACTATAA	1072			
275	ATTAGAAGTAAAGTAGTAGCTTTTCAAGCTTTAAATTTGTAGAACATCTAACTATAA	216			
1073	GTAATATTCGCTGTGCTATCAATATATCTGAGATGTATTCATCTAAAGCATTTTTC	1132			
215	GTAATATTCGCTGTGCTATCAATATATCTGAGATGTATTCATCTAAAGCATTTTTC	156			
1133	ATATCTCAACTAGATTAACCTTTTAGCACAATGCTTAATATCAAGCAGTTGTCATTTGGA	1192			

QY	DB	AGTCACCTGTAAGATGATGTCGCAAGGGAGACACATATTGGATGATATGTTACCATATGT	1252
QY	1193	AGTCACCTGTAAGATGATGTCGCAAGGGAGACACATATTGGATGATATGTTACCATATGT	1252
DB	95	AGTCACCTGTAAGATGATGTCGCAAGGGAGACACATATTGGATGATATGTTACCATATGT	36
QY	1253	TAGGAATAAATATATTTTGGCTG	1275
DB	35	TAGGAATAAATATATTTTGGCTG	13
RESULT	43		
LOCUS	BG777846	624 bp	EST
DEFINITION	602664958F1 NIH_MGC_60	Human sapiens cDNA clone IMAGE:4804574	5'
ACCESSION	BG777846		
VERSION	BG777846.1	GI:14048163	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
TITLE	NIH-MGC http://mgs.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: DCTD/DTF CDNA Library Preparation: CLONETECH Laboratories, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: L10C1651 row: d column: 15 High quality sequence stop: 624.		
FEATURES	Location/Qualifiers		
source	1. 624		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4804574"		
	/clone_lib="NIH_MGC_60"		
	/tissue_type="adenoecarcinoma"		
	/lab_host="DH10B (T1 phage-resistant)"		
	/note="Organ: prostate; Vector: pBR-LIB (Clontech); Site_1: SfiI (ggcgctcgccg); Site_2: SfiI (ggcgctatgccc-3' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATTCTAGAGCCGACGGCGCCGACACGTC(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC library."		
BASE COUNT	213 a 95 c 152 g 164 t		
ORIGIN			
Query Match	39.5%; Score 503; DB 11; Length 624;		
Best Local Similarity	99.8%; Pred. No. 1e-248;		
Matches	623; Conservative 0; Mismatches 0; Indels 1; Gaps 1;		
QY	200	TATGCAGATAGAGATGTTTCACATGTAAGTGAAGTACACCGCTGGTCTATATCAAAA	259
DB	1	TATGCAGATAGAGATGTTTCACATGTAAGTGAAGTACACCGCTGGTCTATATCAAAA	60
QY	260	ATGAGATTTGTCACAAAGGACCTTTAAACCTATTAAACAGAGTGTGAAAAAGGAAA	319

Db 61 ATGAGATTGCTACAAAGGACCCCTTTAAACCTATTAACAAGATGTGAAAAAGGAAA 120
QY 320 CTTCGCTATGTTGGGAATTTGTTCCGTAATAAGATATATCTGGAAGTATGTCATC 379
Db 121 CTTCGCTATGTTGGGAATTTGTTCCGTAATAAGATATATCTGGAAGTATGTCATC 180
QY 380 CCTGAGACTTGGGAAGACCCAGGCGACAAATGATTAACACTGCTGTTGTGTGACAA 439
Db 181 CCTGAGACTTGGGAAGACCCAGGCGACAAATGATTAACACTGCTGTTGTGTGACAA 240
QY 440 GACCAATGATGTTGTTGGAATTTGGAAGTATGTAAGGATGTAAGGATGTAATATG 499
Db 241 GACCAATGATGTTGTTGGAATTTGGAAGTATGTAAGGATGTAAGGATGTAATATG 300
QY 500 GTGAAGTCTTGAAGCATTTGCTGTAATGATGTAAGGATGTAAGGATGTAATATG 558
Db 301 GTGAAGTCTTGAAGCATTTGCTGTAATGATGTAAGGATGTAAGGATGTAATATG 360
QY 559 TGCCATTATGTTGATGATGCTGATGCAAGCAATTAATGATGATGATGCAAG 618
Db 361 TGCCATTATGTTGATGATGCTGATGCAAGCAATTAATGATGATGATGCAAG 420
QY 619 GCTGAACCTGCTACTAGAAAGCTACTGCTGCTGTTGAGAGTATGAGTCTGCA 678
Db 421 GCTGAACCTGCTACTAGAAAGCTACTGCTGCTGTTGAGAGTATGAGTCTGCA 480
QY 679 TGAAGAACCAAGAAATGATGTTGCTGTAATGATGTAAGTATGTAAGTATGCA 738
Db 481 TGAAGAACCAAGAAATGATGTTGCTGTAATGATGTAAGTATGTAAGTATGCA 540
QY 739 TGAATATTTAAAGACCTAGACATGACATTTGGAAGCATTTAGAGTAAAGCAATG 798
Db 541 TGATATTTAAAGACCTAGACATGACATTTGGAAGCATTTAGAGTAAAGCAATG 600
QY 799 AAAAGGATCAGTTCATGATAC 822
Db 601 AAAAGGATCAGTTCATGATAC 624

RESULT 44
LOCUS B1260339 804 bp mRNA EST 17-JUL-2001
DEFINITION 602969372P1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5108993 5',
mRNA sequence.
ACCESSION B1260339
VERSION B1260339.1 GI:14818546
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 804)
TITLE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: L1M11263 row: P column: 18
High quality sequence stop: 737.
Location/Qualifiers
1..804

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5108993"
/clone_id="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"

/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-Sport6; site_1: NotI;
site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.4 kb. Library prepared by Life
Technologies."
BASE COUNT 251 a 147 c 200 g 206 t
ORIGIN

Query Match 39.5%; Score 503; DB 11; Length 804;
Best Local Similarity 99.8%; Pred. No. 1.1e-248;
Matches 553; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 75 AGGACTCCGGGCACTATGAGCGGCTTCAGCAGGAGAGCGCGCGGCTTCGCTG 134
Db 35 AGGACTCCGGGCACTATGAGCGGCTTCAGCAGGAGAGCGCGCGGCTTCGCTG 94
QY 135 ACTACCGAGTCTTCTCAAAAATGAGAAAGCAATATATCTCATTTTCATGATTC 194
Db 95 ACTACCGAGTCTTCTCAAAAATGAGAAAGCAATATATCTCATTTTCATGATTC 154
QY 195 CAATTATGCAATAGATGATGTTTACATGATGATGATGATGATGATGATGATG 254
Db 155 CAATTATGCAATAGATGATGTTTACATGATGATGATGATGATGATGATGATG 214
QY 255 CAAAATGAGATGCTCTACAAAGACCCCTTTAAACCTTTAAACCAAGATGTAAGAA 314
Db 215 CAAAATGAGATGCTCTACAAAGACCCCTTTAAACCTTTAAACCAAGATGTAAGAA 274
QY 315 GAAAATCTGCTATGTTGGGAATTTGTTCCGTAATAAGATATATCTGAACTATG 374
Db 275 GAAAATCTGCTATGTTGGGAATTTGTTCCGTAATAAGATATATCTGAACTATG 334
QY 375 CAAATCTGCTATGTTGGGAATTTGTTCCGTAATAAGATATATCTGAACTATG 434
Db 335 CAAATCTGCTATGTTGGGAATTTGTTCCGTAATAAGATATATCTGAACTATG 394
QY 435 ACAATGACCAATGATGATGTTGTAAGTGAAGCAAGGATGATGCAAGGATGTAAT 494
Db 395 ACAATGACCAATGATGATGTTGTAAGTGAAGCAAGGATGATGCAAGGATGTAAT 454
QY 495 TTGGCGTGAAGTCTTACGATATGTTGCTATATGATGATGATGATGATGATG 554
Db 455 TTGGCGTGAAGTCTTACGATATGTTGCTATATGATGATGATGATGATGATG 514
QY 555 TCATTGCCATTTATGATGATGATGATGATGATGATGATGATGATGATGATG 614
Db 515 TCATTGCCATTTATGATGATGATGATGATGATGATGATGATGATGATGATG 574
QY 615 AACGGCTGAACCT 628
Db 575 AACGGCTGAACCT 588

RESULT 45
LOCUS A1922084/c 624 bp mRNA EST 08-MAR-2000
DEFINITION w88b03.x1 NCI-CCAP_U1 Homo sapiens cDNA clone IMAGE:2452877 3',
similar to SW:IPVR_BOVIN P37980 INORGANIC PYROPHOSPHATASE 1; mRNA
sequence.
ACCESSION A1922084
VERSION A1922084.1 GI:5658048
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 624)
TITLE NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2002, 19:53:05 ; Search time 1849.16 Seconds

(without alignments)
11374.843 Million cell updates/sec

Title: US-09-415-540-2
Perfect score: 1275
Sequence: 1 CAAAGAGTTGGGCTCTCT.....CAAAATTAATATTATTTGCTG 1275

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_hgo_hum:*
31: em_hgo_inv:*
32: em_hgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1275	100.0	1275	6	AR061907	AR061907 Sequence
2	1275	100.0	1275	6	AR084921	AR084921 Sequence
3	1265	99.2	1301	6	BC001022	BC001022 Homo sapi
4	1265	99.2	1303	6	AX015063	AX015063 Sequence
5	1264	99.1	1282	9	AF119665	AF119665 Homo sapi
6	1217	95.5	1239	9	AB026723	AB026723 Homo sapi
7	1185	92.9	1204	9	AF217186	AF217186 Homo sapi
8	1146	89.9	1243	9	AF154065	AF154065 Homo sapi
9	1055	82.7	1200	6	AX018061	AX018061 Sequence
10	788	61.5	856	6	AF108211	AF108211 Homo sapi
11	784	61.5	578	6	AX192855	AX192855 Sequence
12	415	32.5	578	6	AL355138	AL355138 Homo sapi
13	350	27.5	165756	2	AL355138	AL355138 Homo sapi
14	350	27.5	215049	2	AC067749	AC067749 Homo sapi
15	202	15.8	357	6	AX014834	AX014834 Sequence
16	158	12.4	167012	2	AC008429	AC008429 Homo sapi
17	158	12.4	213025	2	AC021159	AC021159 Homo sapi
18	130	10.2	205656	2	AC067966	AC067966 Homo sapi
19	92	7.2	195126	2	AC067746	AC067746 Homo sapi
20	92	7.2	215049	2	AC067749	AC067749 Homo sapi
21	79	6.2	155645	2	AC011012	AC011012 Homo sapi
22	74	5.8	70575	2	AC025103	AC025103 Homo sapi
23	56	4.4	1266	4	BOV10PPP	M95283 Bovine inor
24	50	3.9	1243	10	BC010468	BC010468 Mus muscu
25	44	3.5	70575	2	AC025103	AC025103 Homo sapi
26	25	2.0	144177	9	AL513487	AL513487 Human DNA
27	22	1.7	61946	9	AF020802	AF020802 Homo sapi
28	22	1.7	184180	9	HS1042K10	AL022238 Human DNA
29	22	1.7	216425	2	AC084825	AC084825 Mus muscu
30	22	1.7	323710	2	AC073945	AC073945 Mus muscu
31	22	1.7	323000	9	HS2290A1	AP001724 Homo sapi
32	22	1.7	340000	9	AP001724	AP001724 Homo sapi
33	21	1.6	958	11	CNS061A0	AL400030 TV end of
34	21	1.6	39525	1	SCF91	AL132973 Streptomy
35	21	1.6	44548	9	AF020801	AF020801 Homo sapi
36	21	1.6	51803	9	AF176815	AF176815 Homo sapi
37	21	1.6	60159	9	AL355504	AL355504 Human DNA
38	21	1.6	99461	8	ATP4B14	AL031986 Arabidops
39	21	1.6	106007	8	ATT19K4	AL022373 Arabidops
40	21	1.6	120134	2	AC008034	AC008034 Homo sapi
41	21	1.6	123224	9	AC011414	AC011414 Homo sapi
42	21	1.6	143661	9	AC008491	AC008491 Homo sapi
43	21	1.6	146810	9	HS179115A	292540 Human DNA s
44	21	1.6	155313	2	AC068315	AC068315 Homo sapi
45	21	1.6	157007	2	AC066599	AC066599 Homo sapi

ALIGNMENTS

RESULT	1	PAT	29-SEP-1999
AR061907	AR061907. 1275 bp	DNA	
LOCUS	Sequence 2 from patent US 5843665.		
DEFINITION	AR061907		
ACCESSION	AR061907.1		
VERSION	GI:5989598		
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 1275)		
AUTHORS	Hawkins,P.R. and Hillman,J.L.		
TITLE	Human pyrophosphatase		
JOURNAL	Patent: US 5843665-A 2 01-DEC-1998;		
FEATURES	Location/Qualifiers		
source	1..1275		
BASE COUNT	394 a 235 c 294 g 351 t		1 others
ORIGIN			


```

421 TGCGTGTGTTGTCACATGACCAATGATGTCGTAATGGAGCAAGTATGTC 480
421 TGGCTGTGTTGTCACATGACCAATGATGTCGTAATGGAGCAAGTATGTC 480
481 AAGAGTGAATATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTC 540
481 AAGAGTGAATATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTC 540
541 AAGAGTGAATATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTC 600
541 AAGAGTGAATATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTC 600
601 TATCAATGATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 660
601 TATCAATGATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 660
661 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 720
661 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 720
721 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 780
721 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 780
781 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 840
781 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 840
841 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 900
841 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 900
901 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 960
901 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 960
961 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 1020
961 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 1020
1021 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 1080
1021 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 1080
1081 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 1140
1081 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 1140
1141 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 1200
1141 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 1200
1201 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 1260
1201 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 1260
1261 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 1320
1261 AAGTATATGTCGTAATGTCGTAATGTCGTAATGTCGTAATGTCGTAAT 1320

```

```

REFERENCE
1 (bases 1 to 1301)
AUTHORS
Strausberg, R.
TITLE
Direct Submission
JOURNAL
Submitted (17-NOV-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
MGC project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LNU)
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nih.gov
Shevchenko, Y., Wetherby, R.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, O.L., Mastaglio, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantilip, S., Thomas, P.J.,
Tsongson, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/HLN at: http://image.llnl.gov
Series: IRAL Plate: 5 Row: n Column: 19
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gl: 6563255.
FEATURES
source
1. 1301
/organism="Homo sapiens"
/db_xref="locusid:5464"
/db_xref="taxon:9606"
/clone="MGC:1221 IMAGE:3345496"
/tissue_type="Lymph. Burkitt lymphoma"
/clone_1lb="NIH_MGC_8"
/lab_host="DH10B-R"
/note="Vector: pOT87"
92. 961
/codon_start=1
/product="pyrophosphatase (inorganic)"
/protein_id="AA01022.1"
/db_xref="GI:12654395"
/translaton="MSGSTEERAPFSLRYFLNKEGOVISPFDIPYADKDF
HMYEVRPMNARMEIATKDPIDKIDYKGRKRYVANLEPKYIWNAGAPQME
DPGHNKHTSCGDDNDIPDYCEISRVCAGETIIGVLEILAMIDEGEDMKVIAID
VDDPDAKNDINDIVKRIKRYLEATYVDMRERYKVDGKRENFARNAEKDPAID
IKSTHDMKALVTKTKTKNGKIGISCMNTTISESPKCDPDAARAIVDALPPCESACTV
PTVDKMFHOKN"
BASE COUNT
416 a 237 c 298 g 350 t
ORIGIN
Query Match 99.28; Score 1265; DB 9; Length 1301;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1265; Conservative 0; Mismatches 0; Indels 0; Caps 0;

```

191 ATTCATTTATGAGATAGATGTGTTTCATGATGATGAGTACCAAGCTGCT 250-
194 ATTCATTTATGAGATAGATGTGTTTCATGATGATGAGTACCAAGCTGCT 253
251 AATGCAAAATGAGATGCTACAAAGACCCCTTTAAACCTATTAAACAGATGTGAA 310
254 AATGCAAAATGAGATGCTACAAAGACCCCTTTAAACCTATTAAACAGATGTGAA 313
311 AAGGAAAACTGCTATGCTGCAATTTGTTCCGTTAAAGATATATCTGGAACAT 370
314 AAGGAAAACTGCTATGCTGCAATTTGTTCCGTTAAAGATATATCTGGAACAT 373
371 GGTGCATCCCTCAGACTTGGGAAGACCCAGGCAATGATTAACATAGTGGCTGT 430
374 GGTGCATCCCTCAGACTTGGGAAGACCCAGGCAATGATTAACATAGTGGCTGT 433
431 GGTGCAATGACCAATGATGCTGCAATTTGGAAGCAAGATGTCGAAGAGGTAA 490
434 GGTGCAATGACCAATGATGCTGCAATTTGGAAGCAAGATGTCGAAGAGGTAA 493
491 AATATGCGTGAAGATCTAGGCAATTTGGCTATGATGAGGAAGGGAACCGACTGG 550
494 AATATGCGTGAAGATCTAGGCAATTTGGCTATGATGAGGAAGGGAACCGACTGG 553
551 AATGCAATGCAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 610
554 AATGCAATGCAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 613
611 GTCGAAGGCTGAAGACCGGCTACTAGAGTACTGAGTGGTGGTGGTGGTGGTGG 670
614 GTCGAAGGCTGAAGACCGGCTACTAGAGTACTGAGTGGTGGTGGTGGTGGTGG 673
671 GTTCTGATGGAAGAACCGAAGATGTTGGCTTTAATGCAAGATTTAAAGATAGAC 730
674 GTTCTGATGGAAGAACCGAAGATGTTGGCTTTAATGCAAGATTTAAAGATAGAC 733
731 TTTGCAATGATATTTAAAGACCTATGACCATTTGGAAGCAATTTAGTACTAGAA 790
734 TTTGCAATGATATTTAAAGACCTATGACCATTTGGAAGCAATTTAGTACTAGAA 793
791 ACGAATGGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 850
794 ACGAATGGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 853
851 GATCTGATGCTGCGAGAGCCATTTGATGATGATGATGATGATGATGATGATGAT 910
854 GATCTGATGCTGCGAGAGCCATTTGATGATGATGATGATGATGATGATGATGAT 913
911 ACGATGCAACAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 970
914 ACGATGCAACAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 973
971 GGAATACAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1030
974 GGAATACAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1033
1031 ACGTTTCAACCTTTAAATTTGTAAGTCACTGATGATGATGATGATGATGATGAT 1090
1034 ACGTTTCAACCTTTAAATTTGTAAGTCACTGATGATGATGATGATGATGATGAT 1093
1091 AATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1150
1094 AATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1153
1151 CTTTATGACATGCTTAATATCAAGAGCTGCTATTTGGAAGTCACTTGTGATGAT 1210
1154 CTTTATGACATGCTTAATATCAAGAGCTGCTATTTGGAAGTCACTTGTGATGAT 1213
1211 GTGCAAGGAGACATATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1270
1214 GTGCAAGGAGACATATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1273
1271 TGCTG 1275

Db 1274 TGCTG 1278
RESULT 4
AX015063 1303 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 272 from Patent WO9953040.
ACCESSION AX015063
VERSION AX015063.1 GI:10041202
KEYWORDS
SOURCE human:
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1303)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pilarczyk,C.
TITLE Human nucleic acid sequences from ovarian tumour tissue
JOURNAL Patent: WO 9953040-A 272 21-OCT-1999;
SCHEMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
1. 1303
Location/Dualifiers
source
BASE COUNT 410 a 240 c 300 g 353 t
ORIGIN
Query Match 99.2%; Score 1265; DB 6; Length 1303;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
11 GGGGCTCTCTCTCTGTCAGTGGGGCGCGGCGGGGCTGTGGGCGGCGG 70
17 GGGGCTCTCTCTCTGTCAGTGGGGCGCGGCGGGGCTGTGGGCGGCGG 76
71 GGGGAGAGCTCGGAGCATATGAGCGGCTTCAGCACCGAGAGCGCGCGCTTC 130
77 GGGGAGAGCTCGGAGCATATGAGCGGCTTCAGCACCGAGAGCGCGCGCTTC 136
131 CTGAGATCCGAGCTCTCTCAAAAATGAGAGACATATATTCCTCATTTATGAT 190
137 CTGAGATCCGAGCTCTCTCAAAAATGAGAGACATATATTCCTCATTTATGAT 196
191 ATTCGAATTTATGAGATAGAGATGTTTTCATGATGATGATGATGATGATGATGAT 250
197 ATTCGAATTTATGAGATAGAGATGTTTTCATGATGATGATGATGATGATGATGAT 256
251 AATGCAAAATGAGATGCTACAAAGACCCCTTTAAACCTATTAAACAGATGTGAA 310
257 AATGCAAAATGAGATGCTACAAAGACCCCTTTAAACCTATTAAACAGATGTGAA 316
311 AAGGAAAACTGCTATGCTGCAATTTGTTCCGTTAAAGATATATCTGGAACAT 370
317 AAGGAAAACTGCTATGCTGCAATTTGTTCCGTTAAAGATATATCTGGAACAT 376
371 GGTGCATCCCTCAGACTTGGGAAGACCCAGGCAATGATTAACATAGTGGCTGT 430
377 GGTGCATCCCTCAGACTTGGGAAGACCCAGGCAATGATTAACATAGTGGCTGT 436
431 GGTGCAATGACCAATGATGCTGCAATTTGGAAGCAAGATGTCGAAGAGGTAA 490
437 GGTGCAATGACCAATGATGCTGCAATTTGGAAGCAAGATGTCGAAGAGGTAA 496
491 AATATGCGTGAAGATCTAGGCAATTTGGCTATGATGAGGAAGGGAACCGACTGG 550
497 AATATGCGTGAAGATCTAGGCAATTTGGCTATGATGAGGAAGGGAACCGACTGG 556
551 AATGCAATGCAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 610

Query Match 95.5%; Score 1217; DB 9; Length 1239;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1217; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

BASE COUNT 407 a 220 c 273 g 339 t

ORIGIN

1/gene="SID6-8061"
31..900
/gene="SID6-8061"
/product="Pyrophosphatase"
/catalytic_activity="phosphatase"
/protein_id="BAA84702.1"
/db_xref="GI:5931602"
/translation="MSGFSTEERBAAPFSELEYRVELKNEKGOYISPFHIDIPYADKDFE
HMYVEPRNSAKMEIATKDPINPIKODYKAGKLRVVALFPYKYIINNYA1IPQW
DGNHDKTKCCGGNDPIDYCEIGSKAGKELIGKVLGILAMIDESTOMK1A1N
VDDPDAATYNDINDYKRLKPGYEATVDMERBRKYVDGRENPEFNFNEFDKFAI
IKTSHDKMLVTKTKNGKGISCMNTLTISEPFKDDPDAAIYDALPPEESACTV
PTVDVMEFHOKN"

59 TGGGAGCGGGGGGAGGAGGAGACTCCGGCACTATGAGCGGCTTCACACCAGAGCGCGCC 118
Db 1 TGGGAGCGGGGGGAGGAGGAGACTCCGGCACTATGAGCGGCTTCACACCAGAGCGCGCC 60

119 GCGCCCTTCCTCCCTGGAGTAGCAGAGCTTCCTCAAAAATGAGAAAAGCATAATATATCT 178
Db 61 GCGCCCTTCCTCCCTGGAGTAGCAGAGCTTCCTCAAAAATGAGAAAAGCATAATATATCT 120

179 CCATTCATGATATATCCCAATTTATGAGATTAAGATGTGTTCCATGCTGATGTAAGTA 238
Db 121 CCATTCATGATATATCCCAATTTATGAGATTAAGATGTGTTCCATGCTGATGTAAGTA 180

239 CCAGCGTGTGCTATGCAAAAATGAGATGTGCTACAAAGAGACCTTTAAACCTATTAAA 298
Db 181 CCAGCGTGTGCTATGCAAAAATGAGATGTGCTACAAAGAGACCTTTAAACCTATTAAA 240

299 CAAGATGTGAAAAAAGAAAACTTCGCTATGTGGCAATTTGTTCCTGATTAAGAGATAT 358
Db 241 CAAGATGTGAAAAAAGAAAACTTCGCTATGTGGCAATTTGTTCCTGATTAAGAGATAT 300

359 ATCTGGACCTATGGTGGCACTCCGACAGACTTGGAGAGACCCAGGGCAGCAATGATTAACAT 418
Db 301 ATCTGGACCTATGGTGGCACTCCGACAGACTTGGAGAGACCCAGGGCAGCAATGATTAACAT 360

419 ACTGCTCTTGTGTGCAATGAGCCCAATGATGTGTGTGAAATTTGGAAGCAAGGTATGT 478
Db 361 ACTGCTCTTGTGTGCAATGAGCCCAATGATGTGTGTGAAATTTGGAAGCAAGGTATGT 420

479 GCAAGAGGTGAATTAATTTGGCGTGAAGTCTTAGGATATTTGGCTATGTTATGCCAAGG 538
Db 421 GCAAGAGGTGAATTAATTTGGCGTGAAGTCTTAGGATATTTGGCTATGTTATGCCAAGG 480

539 GAAACCGCATGGAAGTCAATTTGCCATTAAATGTGGATGATCTCGATGACGACCAATTATAT 598
Db 481 GAAACCGCATGGAAGTCAATTTGCCATTAAATGTGGATGATCTCGATGACGACCAATTATAT 540

599 GATATCATGATGTCAAAACGGCTGAACACCGGTGACTTAAAGCTACTGTGACGAGCTTT 658
Db 541 GATATCAAGATGTCAAAACGGCTGAACACCGGTGACTTAAAGCTACTGTGACGAGCTTT 600

659 AGAAGGTATAAGGTTCTGATGGAAGAACCGAAAAATGATGTTTGGCTTTAATGCAGATTT 718
Db 601 AGAAGGTATAAGGTTCTGATGGAAGAACCGAAAAATGATGTTTGGCTTTAATGCAGATTT 660

719 AAGATTAAGGACTTTGCCATTGATTAATTAAGGACCTGATACCATTTGGAAGCATTA 778
Db 661 AAGATTAAGGACTTTGCCATTGATTAATTAAGGACCTGATACCATTTGGAAGCATTA 720

QY 839 CCCTTCAGTGTGATCCTGATGCTGCCAGAGCAGCTTGTGATGCTTTTACACCACCCCTCT 898
 DB 781 CCTTCAAGTGTGATGCTGATGCTGCCAGAGCAGCTTGTGATGCTTTTACACCACCCCTCT 840
 QY 899 GAATGTGCTGCACAGTACACAGACGCTGATGCTTGTGATGCTTTTACACCACCCCTCT 958
 DB 841 GAATGTGCTGCACAGTACACAGACGCTGATGCTTGTGATGCTTTTACACCACCCCTCT 900
 QY 959 TGAGATTTCTGCGAATACAGCTATGCTTGTGATGCTTTTACACCACCCCTCT 1018
 DB 901 TGAGATTTCTGCGAATACAGCTATGCTTGTGATGCTTTTACACCACCCCTCT 960
 QY 1019 AGTAAAGTAGTGTGCTTTTCAAGCTTTTAAATTTGTAGAACTCATCTAACTAAAGTAAAT 1078
 DB 961 AGTAAAGTAGTGTGCTTTTCAAGCTTTTAAATTTGTAGAACTCATCTAACTAAAGTAAAT 1020
 QY 1079 TCTGCTGTGACTATATCATATATGCTGATGCTTTTACACCACCCCTCT 1138
 DB 1021 TCTGCTGTGACTATATCATATATGCTGATGCTTTTACACCACCCCTCT 1080
 QY 1139 CAAGTACAGTGTGCTTTTCAAGCTTTTAAATTTGTAGAACTCATCTAACTAAAGTAAAT 1198
 DB 1081 CAAGTACAGTGTGCTTTTCAAGCTTTTAAATTTGTAGAACTCATCTAACTAAAGTAAAT 1140
 QY 1199 TTGTGAATAGATGTGCAAGGGGAGCAGATTTGTGATGCTTTTACACCACCCCTCT 1258
 DB 1141 TTGTGAATAGATGTGCAAGGGGAGCAGATTTGTGATGCTTTTACACCACCCCTCT 1200
 QY 1259 ATAAATTTATTTGCTG 1275
 DB 1201 ATAAATTTATTTGCTG 1217

RESULT 7

AF217186 1204 bp mRNA PRI -02-DEC-2000
 LOCUS AF217186 Homo sapiens inorganic pyrophosphatase 1 (PPAI) mRNA, complete cds.
 DEFINITION AF217186
 VERSION AF217186.1 GI:11526786
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 AUTHORS Kannil, L., Johansson, M. and Karlsson, A.
 TITLE Cloning of a human inorganic pyrophosphatase cDNA
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1204)
 AUTHORS Kannil, L., Johansson, M. and Karlsson, A.
 TITLE Direct Submission
 JOURNAL Submitted (16-DEC-1999) Clinical Virology, Karolinska Institute,
 Huddinge University Hospital, Stockholm S-14186, Sweden
 FEATURES
 source
 1..1204
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..1204
 /gene="PPAI"
 1..870
 /gene="PPAI"
 /note="phase-1"
 /codon_start=1
 /product="inorganic pyrophosphatase 1"
 /protein_id="AAG36780.1"
 /db_xref="GI:11526786"
 /translation="MSGFSTERAPLESLEYRVFLKNEKGYISPHDIPYADKOV
 HMYVEPRMSNAKMEIATKDLNPIKDVKGKGLRVANLEPPYKGIINNYGAIPTWE
 DPNDKHTCCGDDPIIDVCEIGSCVAREIIGVGLIAMIIDESETDMKVIIN
 VDDPDANYNDINDYKRLKPGYLEATVMPFRYKVPDOKPEPEFAPNPEFMDKDAID
 IIKSTHDKMLATLVKTKNGKISCHNTLLSSPFCDDAARAIYDALPPCESACTV
 PTDVKKRHHOKN"

BASE COUNT 403 a 210 c 258 g 333 t

ORIGIN

Query Match 92.9%: Score 1185; DB 9; Length 1204;
 Best Local Similarity 100.0%: Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
 Matches 1185: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 ATGAGCGGCTTGACGACCCAGGAGGCGCGCGCCCTTCCTCGAGTACCGAGTCTTC 148
 DB 1 ATGAGCGGCTTGACGACCCAGGAGGCGCGCGCCCTTCCTCGAGTACCGAGTCTTC 60
 QY 149 CTCAAAATGAGAAAAGACAAATATATATCTCATTTTCATGATATTCATATGAGAT 208
 DB 61 CTCAAAATGAGAAAAGACAAATATATATCTCATTTTCATGATATTCATATGAGAT 120
 QY 209 AAGGATGTGTTTACATGTTAGTGTGAGTACAGCGTGGTCTTATATCCAAAATGAGATT 268
 DB 121 AAGGATGTGTTTACATGTTAGTGTGAGTACAGCGTGGTCTTATATGCAAAAATGAGATT 180
 QY 269 GCTACAAAGGACCCCTTAAACCTTATTAACAAGATGTGAAAAAGAAAACCTTCCCTAT 328
 DB 181 GCTACAAAGGACCCCTTAAACCTTATTAACAAGATGTGAAAAAGAAAACCTTCCCTAT 240
 QY 329 GTTGGCAATTTGTTCCGATATAAGGATATATCTGGAACATATGTCCTCAGACT 388
 DB 241 GTTGGCAATTTGTTCCGATATAAGGATATATCTGGAACATATGTCCTCAGACT 300
 QY 389 TGGGAAGACCCAGGCGCAATGATATAACATCTGCTGTTGTGTGACAAATGCCCAATT 448
 DB 301 TGGGAAGACCCAGGCGCAATGATATAACATCTGCTGTTGTGTGACAAATGCCCAATT 360
 QY 449 GATGTGTGTAATTTGGAAGCAAGATATGTCGAAGAGTGAATAATTTGGCGTGAAGTT 508
 DB 361 GATGTGTGTAATTTGGAAGCAAGATATGTCGAAGAGTGAATAATTTGGCGTGAAGTT 420
 QY 509 CTAGCATATTTGGCTATGATTTGACGAGGGAACCCGACTGGAATCATTTGCCATTAT 568
 DB 421 CTAGCATATTTGGCTATGATTTGACGAGGGAACCCGACTGGAATCATTTGCCATTAT 480
 QY 569 GTGATGATCTCGATGACCCCAATTTATGATATCAATGATGTCGAAAGGCGTGAACCT 628
 DB 481 GTGATGATCTCGATGACCCCAATTTATGATATCAATGATGTCGAAAGGCGTGAACCT 540
 QY 629 GCTACTTGAAGCTCTGCTGAGCTGTTTGAAGATATAGGTTCTGATGGAAGAACCA 688
 DB 541 GCTACTTGAAGCTCTGCTGAGCTGTTTGAAGATATAGGTTCTGATGGAAGAACCA 600
 QY 689 GAAATGAGTTTGCCTTTATGACGAGATTTAAAGATAGGACTTTGCCATTGATATT 748
 DB 601 GAAATGAGTTTGCCTTTATGACGAGATTTAAAGATAGGACTTTGCCATTGATATT 660
 QY 749 AAAGGACATCATGACCATTTGGAAGCATTTAGTACCTAAGAAAAGCAATGGAAGAAATC 808
 DB 661 AAAGGACATCATGACCATTTGGAAGCATTTAGTACCTAAGAAAAGCAATGGAAGAAATC 720
 QY 809 AGTGCATGAATATACACTTTGCTGAGAGCCCTTAAAGTGTGATCTGATGCTGCAGAA 868
 DB 721 AGTGCATGAATATACACTTTGCTGAGAGCCCTTAAAGTGTGATCTGATGCTGCAGAA 780
 QY 869 GCCATTGTGATGCTTTTACACACCCCTGTGAATCTGCTGCACAGTACCAACAGAGCTG 928
 DB 781 GCCATTGTGATGCTTTTACACACCCCTGTGAATCTGCTGCACAGTACCAACAGAGCTG 840
 QY 929 GATTAAGTGTGATGCTTTTACACACCCCTGTGAATCTGCTGCACAGTACCAACAGAGCTG 988
 DB 841 GATTAAGTGTGATGCTTTTACACACCCCTGTGAATCTGCTGCACAGTACCAACAGAGCTG 900
 QY 989 GCTACATCTGTTTCATCTGATGATATTTAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1048
 DB 901 GCTACATCTGTTTCATCTGATGATATTTAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 960
 QY 1049 ATTTGTAGACTCATCTAACTAAAGTAAATTTGCTGTGACTATATCAATATATCTAGAA 1108
 DB 960 ATTTGTAGACTCATCTAACTAAAGTAAATTTGCTGTGACTATATCAATATATCTAGAA 880

Db 961 ATTGTGAGACTCATCTACTAAGTAATTCGTGCTGACTAATCCATATACTCAGAA 1020
OY 1109 TGTATCATCTAAGACATTTTTCATATCTCACTAGATACCTTTAGACATGCTTAA 1168
Db 1021 TGTATCATCTAAGACATTTTTCATATCTCACTAGATACCTTTAGACATGCTTAA 1080
OY 1169 ATATCAAGACGTTGTCTATTTGGAAGTCACCTTGTGAATAGATGTGCAAGGGGAGACATA 1228
Db 1081 ATATCAAGACGTTGTCTATTTGGAAGTCACCTTGTGAATAGATGTGCAAGGGGAGACATA 1140
OY 1229 TTGGATGTAATGTGTACCATATGTGTAGAAATTAATTTATTTGTC 1273
Db 1141 TTGGATGTAATGTGTACCATATGTGTAGAAATTAATTTATTTGTC 1185

RESULT 8
AF154065 1243 bp mRNA PRI 09-NOV-1999
LOCUS AF154065 Homo sapiens inorganic pyrophosphatase (pp) mRNA, complete cds.
DEFINITION AF154065
ACCESSION AF154065
VERSION AF154065.1 GI:4960207
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1243)
Fairchild, T.A. and Patejunas, G.
Cloning and expression profile of human inorganic pyrophosphatase
JOURNAL Biochim. Biophys. Acta 1447 (2-3), 133-136 (1999)
MEDLINE 20011271
REFERENCE 2 (bases 1 to 1243)
Patejunas, G.
Direct Submission
Submitted (25-MAY-1999) Molecular Cardiology Laboratory, Evanston
Northwestern Healthcare Research Institute, 2650 Ridge Avenue,
Evanston, IL 60201, USA
FEATURES
source location/Qualifiers
1..1243
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10q21.3"
/tissue.type="heart"
1..1243
/gene="pp"
46..915
/gene="pp"
/EC_number="3.6.1.1"
/function="catalyzes hydrolysis of pyrophosphate"
/codon.start=1
/product="inorganic pyrophosphatase"
/protein_id="A03463.1"
/db_xref="GI:4960208"
/translation="MSGSTEERAPDSLELYFLNKGQSYSPHDIPIYADKDF
HMYVEYRMSNAKMEIATDPLNPKIDYKKGKLRVANLEPYKGIWNGAIPOHWE
DEBHNKHTGGCCSDNDIDVCEIYSKVGARGLIGKVGITLAMDGEETDMKVIAIN
YDDPDANNDINDYKRLRPGIYEATVDMFRKRYVDGRENFAFAAEKKDFRID
ILKSTHMKALVTKTKNGKISCMNTTISEPFCDDPDARAIVDALPPECSEACTV
PTVDVRFHFKHOKN"

BASE COUNT 400 a 224 c 279 g 340 t
ORIGIN

Query Match 89.9%; Score 1146; DB 9; Length 1243;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1196; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 79 CTCGCGCATATGAGCGCTTCAGACCGAGGAGCGCGCGCCCTTCCTCGAGTA 138
Db 36 CTCGCGCATATGAGCGCTTCAGACCGAGGAGCGCGCGCCCTTCCTCGAGTA 95
OY 139 CCGAGTCTTCTCTCAAAATGAGAAAGACAATATATCTCATTTCCATATTTCCAAAT 198

Db 96 CCGAGTCTTCTCTCAAAATGAGAAAGACAATATATATCTCATTTTCAATATTTCCAAAT 155
OY 199 TTATGACAGATAGAGATGTGTTTACATGTAGTGAAGTACACAGCTGTCTAAATGCAAA 258
Db 156 TTATGACAGATAGAGATGTGTTTACATGTAGTGAAGTACACAGCTGTCTAAATGCAAA 215
OY 259 AATGAGATTTGCTCAAGAGACCCCTTTAAACCCCTATTAAACAAATGTGAAAAAGGAAA 318
Db 216 AATGAGATTTGCTCAAGAGACCCCTTTAAACCCCTATTAAACAAATGTGAAAAAGGAAA 275
OY 319 ACTTCGATATGTCGAAATTTGTTCCCGTATAAGATATATGCAACATGTGTCGCAT 378
Db 276 ACTTCGATATGTCGAAATTTGTTCCCGTATAAGATATATGCAACATGTGTCGCAT 335
OY 439 TGACCCCAATGTATGTGTGTAATTTGGAAGCAAGATATGTGCAGAGAGTGAATTAATTTG 498.
Db 396 TGACCCCAATGTATGTGTGTAATTTGGAAGCAAGATATGTGCAGAGAGTGAATTAATTTG 455
OY 499 CGTGAAGTTCTAGCAATATTTGCTATGATGAGCAAGGGGAAACCGACATGCAATGTCAT 558
Db 456 CGTGAAGTTCTAGCAATATTTGCTATGATGAGCAAGGGGAAACCGACATGCAATGTCAT 515
OY 559 TGCCATTAAATGTGATGATCCTGATGACGCAATTTATATATATCAATGATGCAAAAG 618
Db 516 TGCCATTAAATGTGATGATCCTGATGACGCAATTTATATATATCAATGATGCAAAAG 575
OY 619 GCTGAACCTGTGCTACTTAGAAGTACTGTGACTGTTTGAAGATTAAGATTTGCTCTGA 678
Db 576 GCTGAACCTGTGCTACTTAGAAGTACTGTGACTGTTTGAAGATTAAGATTTGCTCTGA 635
OY 679 TGAACCAACGAAATAGTTTGGCTTAAATGCAATTTAAAGATPAGGACTTGGCAT 738
Db 636 TGAACCAACGAAATAGTTTGGCTTAAATGCAATTTAAAGATPAGGACTTGGCAT 695
OY 739 TGATATTTATTAAGACATGACATGACATGGAAGCAATTAAGTACTAAGAAACGAATGG 798
Db 696 TGATATTTATTAAGACATGACATGACATGGAAGCAATTAAGTACTAAGAAACGAATGG 755
OY 799 AAAAGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 858
Db 756 AAAAGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 815
OY 859 TGCTGCCAGAGCATTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 918
Db 816 TGCTGCCAGAGCATTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 875
OY 919 AACAGACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 978
Db 876 AACAGACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 935
OY 979 AGCGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1038
Db 936 AGCGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 995
OY 1039 AAGGCTTTAATTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1098
Db 996 AAGGCTTTAATTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1055
OY 1099 ATACTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1158
Db 1056 ATACTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1115
OY 1159 ACATGCTTAATATCAAGAGAGTTGATTTGGAAGTACCTTTGATATGATGATGATGATGATGATGAT 1218
Db 1116 ACATGCTTAATATCAAGAGAGTTGATTTGGAAGTACCTTTGATATGATGATGATGATGATGATGAT 1175
OY 1219 GGAGCATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1275

Db 1176 GGAGCACAATATGATATATGTACCATATGTAGGAATAAATTATTGCTG 1232

RESULT 9
AX018061 1200 bp DNA PAT 07-SEP-2000
LOCUS Sequence 3 from Patent W0946374.
DEFINITION AX018061
ACCESSION AX018061
VERSION AX018061.1 GI:10042512
KEYWORDS human.
SOURCE human sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1200)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pillarsky,C.
TITLE Human nucleic acid sequences from prostate tumour tissue
JOURNAL Patent: WO 946374-A 3 16-SEP-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILLARSKY CHRISTIAN (DE)
FEATURES
Location/Qualifiers
1..1200
source /organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 368 a 227 c 278 g 327 t
ORIGIN

Query Match 82.7%; Score 1055; DB 6; Length 1200;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1055; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CCTTCCCTCCGAGTACCGAGTCTCTCTCAAAAATGGAAGACCAATATATCTCCAT 182
DB 111 CCTTCCCTCCGAGTACCGAGTCTCTCTCAAAAATGGAAGACCAATATATCTCCAT 170
QY 183 TTGATGATATTTCCAAATTTATGACAGATAGATGTTTCATCATGTTGAGTACAC 242
DB 171 TTGATGATATTTCCAAATTTATGACAGATAGATGTTTCATCATGTTGAGTACAC 230
QY 243 GCTGTGCTAAATGCAAAAATGAGATTGCTACAAAGACCCCTTAAACCTATTAAACAG 302
DB 231 GCTGTGCTAAATGCAAAAATGAGATTGCTACAAAGACCCCTTAAACCTATTAAACAG 290
QY 303 ATGTGAAAAAGAAAAAAGCTGCTATGTTGCGAATTTGTTCCCTATTAAGATATATCT 362
DB 291 ATGTGAAAAAGAAAAAAGCTGCTATGTTGCGAATTTGTTCCCTATTAAGATATATCT 350
QY 363 GGAACATATGTCATCCCTCAGACTGGGAGAACCCAGGGCACAATGATTAACATCTG 422
DB 351 GGAACATATGTCATCCCTCAGACTGGGAGAACCCAGGGCACAATGATTAACATCTG 410
QY 423 GCTGTGCTGATGCAATGACCCCAATGATGTTGTAATTTGGAAGCAAGTATGTCAA 482
DB 411 GCTGTGCTGATGCAATGACCCCAATGATGTTGTAATTTGGAAGCAAGTATGTCAA 470
QY 483 GAGGTGAATAATTTGGCGTGAAGTTCTAGGCATATTTGGCTATGATGAGAGGGGAAA 542
DB 471 GAGGTGAATAATTTGGCGTGAAGTTCTAGGCATATTTGGCTATGATGAGAGGGGAAA 530
QY 543 CCAGCTGGAAGTATGTCATTAATGATGATGATGATGATGATGATGATGATGATGATG 602
DB 531 CCAGCTGGAAGTATGTCATTAATGATGATGATGATGATGATGATGATGATGATGATG 590
QY 603 TCAATGATGTCAAAGCGCTGAAGCTGCTAGTAAAGTACTGCTGAGCTGTTAGAA 662
DB 591 TCAATGATGTCAAAGCGCTGAAGCTGCTAGTAAAGTACTGCTGAGCTGTTAGAA 650
QY 663 GGTATAGGTTCTGATGAGAAACAGAAATGAGTTGGTTTAATGACAATTTAAAG 722
DB 651 GGTATAGGTTCTGATGAGAAACAGAAATGAGTTGGTTTAATGACAATTTAAAG 710

QY 723 ATAGGACTTTGCCATGATATTTAATAAGCACTCATGACCATTTGGAAGCATTTAGTA 782
DB 711 ATAGGACTTTGCCATGATATTTAATAAGCACTCATGACCATTTGGAAGCATTTAGTA 770
QY 783 CTAGAAAGCAATGAAAAAGATTCAGTTGCATGATTAACAATTTGTTGAGAGCCCT 842
DB 771 CTAGAAAGCAATGAAAAAGATTCAGTTGCATGATTAACAATTTGTTGAGAGCCCT 830
QY 843 TCAAGTGTGATCTGATGCTGCGAGAGCCATTTGATGCTTTACACACCCCTGTAAT 902
DB 831 TCAAGTGTGATCTGATGCTGCGAGAGCCATTTGATGCTTTACACACCCCTGTAAT 890
QY 903 CTGCTCAGACAGTACCAAGAGAGCTGATAGTGTTCATCACCAGAAAAACAATAG 962
DB 891 CTGCTCAGACAGTACCAAGAGAGCTGATAGTGTTCATCACCAGAAAAACAATAG 950
QY 963 ATTCTCTGAAATCAAGCTGATATTTGCTCATGCTGTTCAATGATATTAAGATA 1022
DB 951 ATTCTCTGAAATCAAGCTGATATTTGCTCATGCTGTTCAATGATATTAAGATA 1010
QY 1023 AAATGATGCTTTTCAAGCTTTAATTTGTAGAACTCATCTAAGTAAATTTCTG 1082
DB 1011 AAATGATGCTTTTCAAGCTTTAATTTGTAGAACTCATCTAAGTAAATTTCTG 1070
QY 1083 CTGAGCTAATCCAAATATCTCAGATGTTATTCATCTAAGCAATTTTTCATCTCAAC 1142
DB 1071 CTGAGCTAATCCAAATATCTCAGATGTTATTCATCTAAGCAATTTTTCATCTCAAC 1130
QY 1143 TAAGTAACTTTTGAACATGCTTAATAATATCAAG 1177
DB 1131 TAAGTAACTTTTGAACATGCTTAATAATATCAAG 1165

RESULT 10
AX014862 856 bp DNA PAT 07-SEP-2000
LOCUS Sequence 52 from Patent W0953040.
DEFINITION AX014862
ACCESSION AX014862
VERSION AX014862.1 GI:10041129
KEYWORDS human.
SOURCE human sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 856)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pillarsky,C.
TITLE Human nucleic acid sequences from ovarian tumour tissue
JOURNAL Patent: WO 953040-A 52 21-OCT-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILLARSKY CHRISTIAN (DE)
FEATURES
Location/Qualifiers
1..856
source /organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 251 a 163 c 222 g 220 t
ORIGIN

Query Match 61.8%; Score 788; DB 6; Length 856;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 788; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GGGCTCTCTCTCTGTCAGTCGGCGCGCTGCGGCTGTGCTGTGTCAGCGCGG 70
DB 17 GGGCTCTCTCTCTGTCAGTCGGCGCGCTGCGGCTGTGCTGTGTCAGCGCGG 76
QY 71 CGGAGAGACTCGGCATATAGCGGCTTACGACCGAGAGCGCGCGCCCTTCTCC 130
DB 77 CGGAGAGACTCGGCATATAGCGGCTTACGACCGAGAGCGCGCGCCCTTCTCC 136

OY 131 CTGAGTACGAGCTCTCTCTCAAAATGAGAAAGCAATATATATCTCCATTCTCAT 190
 |||||
 Db 137 CTGAGTACGAGCTCTCTCTCAAAATGAGAAAGCAATATATATCTCCATTCTCAT 196
 OY 191 ATTCCATTATGACATGATGATGTTTCATGATGATGATGATGATGATGATGATGAT 250
 |||||
 Db 197 ATTCCATTATGACATGATGATGTTTCATGATGATGATGATGATGATGATGATGAT 256
 OY 251 AATGCAAAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 310
 |||||
 Db 257 AATGCAAAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 316
 OY 311 AAGGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 370
 |||||
 Db 317 AAGGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 376
 OY 371 GGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 430
 |||||
 Db 377 GGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 436
 OY 431 GGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 490
 |||||
 Db 437 GGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 496
 OY 491 AATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 550
 |||||
 Db 497 AATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 556
 OY 551 AATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 610
 |||||
 Db 557 AATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 616
 OY 611 GTCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 670
 |||||
 Db 617 GTCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 676
 OY 671 GTCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 730
 |||||
 Db 677 GTCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 736
 OY 731 TTTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 790
 |||||
 Db 737 TTTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 796
 OY 791 ACGATG 798
 |||||
 Db 797 ACGATG 804
 |||||
 RESULT 11
 AF108211 846 bp mRNA PRI 14-APR-1999
 LOCUS Homo sapiens cytosolic inorganic pyrophosphatase mRNA, partial cds.
 DEFINITION AF108211
 ACCESSION AF108211
 VERSION AF108211.1 GI:4583152
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Metazoa; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Rumsfeld, J., Ziegelbauer, K. and Spaltmann, F.
 TITLE Cloning, expression, affinity purification and characterization of
 polyhistidine-tagged cytosolic Saccharomyces cerevisiae and human
 inorganic pyrophosphatases for differential screening of compounds
 for antifungal activity
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 846)
 AUTHORS Rumsfeld, J., Ziegelbauer, K. and Spaltmann, F.
 TITLE Direct Submission
 JOURNAL Submitted (20-NOV-1998) Research Antinfectives 1, Bayer AG, P.O.
 Box 10179, Wuppertal 42096, Germany
 FEATURES
 Location/Qualifiers

source 1. 846
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="BjAB"
 /cell_type="B-cell"
 <1. >846
 /note="ipase"
 /codon_start=1
 /product="cytosolic inorganic pyrophosphatase"
 /protein_id="A02496.1"
 /db_xref="GI:4583152"
 /translation="STEERAAFLSLEKRPVPIKREKQVYSPEDIPYADKDFVHNV
 EVPEWSNAKMEIAIKDPLNPKDVPKGLRYANLFPYKGYIMNAGLIPQWEDGH
 NDRKTCGCDNDPIDVEIGSKVCAKEGILGVVLTILMIDBEGTMDKVIDNDP
 DAANYNDINVRKLPGLYATVDWERRRYVPPGKENEFAFAEKRODPAIDIKS
 TMDHKKALVTKTKTGKISCMNTLISSEPKCDPAARAIVDLAPPCSACTVPIDV
 DKWHH"
 BASE COUNT 270 a 155 c 201 g 220 t
 ORIGIN
 Query Match 61.5%; Score 784; DB 9; Length 846;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 784; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 123 CCTTCTCCCTGAGTACGAGCTCTCTCAAAATGAGAAAGCAATATATATCTCCAT 182
 |||||
 Db 23 CCTTCTCCCTGAGTACGAGCTCTCTCAAAATGAGAAAGCAATATATATCTCCAT 82
 OY 183 TTCAATGATATGCAATTTATGCAATGATGATGATGATGATGATGATGATGATGAT 242
 |||||
 Db 83 TTCAATGATATGCAATTTATGCAATGATGATGATGATGATGATGATGATGATGAT 142
 OY 243 GCTGCTCTAATGCAAAATGAGATGCTCAAAAGACCTTTAAACCTATTAAACAAG 302
 |||||
 Db 143 GCTGCTCTAATGCAAAATGAGATGCTCAAAAGACCTTTAAACCTATTAAACAAG 202
 OY 303 ATGTGAAAAAGGAAACCTGCTATGCTGCAATTTGTTCCCTATTAAGATATATCT 362
 |||||
 Db 203 ATGTGAAAAAGGAAACCTGCTATGCTGCAATTTGTTCCCTATTAAGATATATCT 262
 OY 363 GGAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 422
 |||||
 Db 263 GGAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 322
 OY 423 GCTGTTGTGGAATGACCAATGATGATGATGATGATGATGATGATGATGATGATGAT 482
 |||||
 Db 323 GCTGTTGTGGAATGACCAATGATGATGATGATGATGATGATGATGATGATGATGAT 382
 OY 483 GAGTGAATATATGCTGGAAGTCTAGACATATGCTATGATGATGATGATGATGATGAT 542
 |||||
 Db 383 GAGTGAATATATGCTGGAAGTCTAGACATATGCTATGATGATGATGATGATGATGAT 442
 OY 543 CCGACTGGAAGTCAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 602
 |||||
 Db 443 CCGACTGGAAGTCAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 502
 OY 603 TCAATGATGCAAAAGGCTGGAACCTGCTACTTGAAGTCACTGCTGCTGCTGCTGCTGCT 662
 |||||
 Db 503 TCAATGATGCAAAAGGCTGGAACCTGCTACTTGAAGTCACTGCTGCTGCTGCTGCTGCT 562
 OY 663 GGTATAGGTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 722
 |||||
 Db 723 GGTATAGGTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 682
 OY 783 CTAGAAAGCAATGCAAAAGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 842
 |||||
 Db 683 CTAGAAAGCAATGCAAAAGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 742


```
misc_feature 10542..11663
/note="assembly_name:Contig73"
misc_feature 11764..12926
/note="assembly_name:Contig75"
misc_feature 13027..14168
/note="assembly_name:Contig112"
misc_feature 14269..15574
/note="assembly_name:Contig113"
misc_feature 15675..17038
/note="assembly_name:Contig116"
misc_feature 17139..18401
/note="assembly_name:Contig118"
misc_feature 18502..19788
/note="assembly_name:Contig121"
misc_feature 19889..21245
/note="assembly_name:Contig124"
misc_feature 21346..23067
/note="assembly_name:Contig127"
misc_feature 23168..25167
/note="assembly_name:Contig134"
clone_end:SP6"
misc_feature 25268..28266
/note="assembly_name:Contig135"
misc_feature 28367..30737
/note="assembly_name:Contig136"
misc_feature 30838..35830
/note="assembly_name:Contig137"
misc_feature 35931..43873
/note="assembly_name:Contig138"
misc_feature 43974..54313
/note="assembly_name:Contig139"
misc_feature 54414..62524
/note="assembly_name:Contig140"
misc_feature 62625..75784
/note="assembly_name:Contig141"
misc_feature 75885..92445
/note="assembly_name:Contig142"
misc_feature 92546..109803
/note="assembly_name:Contig143"
misc_feature 109904..128494
/note="assembly_name:Contig144"
misc_feature 128595..146739
/note="assembly_name:Contig145"
misc_feature 146840..167230
/note="assembly_name:Contig146"
misc_feature 167331..187365
/note="assembly_name:Contig147"
misc_feature 187466..215049
/note="assembly_name:Contig148"
clone_end:T7"

BASE COUNT 57394 a 47838 c 47126 g 59466 t 3225 others
ORIGIN

Query Match 27.5%; Score 350; DB 2; Length 215049;
Best Local Similarity 100.0%; Pred. No. 1.1e-167;
Matches 350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 926 GTGATAGTGGTTCATCACCAGAAAACTAATGAGATTCTCGAATACAGCTGAT 985
|||||
DB 113781 GTGATAGTGGTTCATCACCAGAAAACTAATGAGATTCTCGAATACAGCTGAT 113840
|||||
QY 986 ATTGCTACATCGTTTCATCTGAGATGATTAAGAGTAAAGTAGTACCTTTCAAGCTT 1045
|||||
DB 113841 ATTGCTACATCGTTTCATCTGAGATGATTAAGAGTAAAGTAGTACCTTTCAAGCTT 113900
|||||
QY 1046 TAAATTTAGAACATCTACTAAGTAATTCGCTGAGTAAATCAATATATACCTA 1105
|||||
DB 113901 TAAATTTAGAACATCTACTAAGTAATTCGCTGAGTAAATCAATATATACCTA 113960
|||||
QY 1106 GAATGTTATCCATCTAAGCAATTTTTCATATCTCACTAAGTAACTTTTACACATGCT 1165
|||||
DB 113961 GAATGTTATCCATCTAAGCAATTTTTCATATCTCACTAAGTAACTTTTACACATGCT 114020
|||||
```

```
QY 1166 TAAATATCAAAACAGTGTTCATTTGGAACTCTGTGAATAGATGCAAGGAGCAC 1225
|||||
DB 114021 TAAATATCAAAACAGTGTTCATTTGGAACTCTGTGAATAGATGCAAGGAGCAC 114080
|||||
QY 1226 AATTGGAGTGTATGTTTACCATATGTTAGGAATTAATTTTTCGCG 1275
|||||
DB 114081 AATTGGAGTGTATGTTTACCATATGTTAGGAATTAATTTTTCGCG 114130
|||||

RESULT 15
AX014834
LOCUS AX014834 357 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 23 from Patent WO953040.
ACCESSION AX014834
VERSION AX014834.1 GI:10041101
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 357)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pillarsky,C.
TITLE Human nucleic acid sequences from ovarian tumour tissue
JOURNAL Patent: WO 953040-A 23 21-OCT-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METZGER GRS FUER GENOMFORSCHUN
(DE); PILARSKI CHRISTIAN (DE)
FEATURES
source 1..357
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 74 a 89 c 96 g 98 t
ORIGIN

Query Match 15.8%; Score 202; DB 6; Length 357;
Best Local Similarity 100.0%; Pred. No. 6.2e-92;
Matches 202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GGGGCTCTCTCTCTGTGCTGCTGCGCGCGCGTGTGCTGCTGCTGCTGCGCGG 70
|||||
DB 17 GGGGCTCTCTCTCTGTGCTGCTGCGCGCGCGTGTGCTGCTGCTGCTGCGCGG 76
|||||
QY 71 CGGAGGACTCCGGGACATATGAGCGGCTTCAGCAGCAGGAGGCGCGCGCTCTCC 130
|||||
DB 77 CGGAGGACTCCGGGACATATGAGCGGCTTCAGCAGCAGGAGGCGCGCGCTCTCC 136
|||||
QY 131 CTGGAGTACCGAGTCTCTCTCAAAAATGAGAAAGCAATATATATCTCATTTATGAT 190
|||||
DB 137 CTGGAGTACCGAGTCTCTCTCAAAAATGAGAAAGCAATATATATCTCATTTATGAT 196
|||||
QY 191 ATTCCAAATTATGACAGATAAGG 212
|||||
DB 197 ATTCCAAATTATGACAGATAAGG 218
|||||

RESULT 16
AC008429
LOCUS AC008429 167012 bp DNA HTG 20-APR-2001
DEFINITION Homo sapiens chromosome 5 clone CTC-308K20, WORKING DRAFT SEQUENCE,
9 unordered pieces.
ACCESSION AC008429
VERSION AC008429.4 GI:13699351
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 167012)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
```

JOURNAL
REFERENCE 2 (bases 1 to 167012)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Apr 20, 2001 this sequence version replaced gi:7708851.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Project Name: 303711
Center clone name: CIT-HSPC_308K20

Summary Statistics
Consensus quality: 156314 bases at least Q40
Consensus quality: 158094 bases at least Q30
Consensus quality: 158732 bases at least Q20
Estimated insert size: 16200; pulse field gel estimation
Estimated insert size: 166212; sum-of-coverage estimation
Quality coverage: 8.74 in Q20 bases; sum-of-coverage estimation
Quality coverage: 8.52 in Q20 bases; sum-of-coverage estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1151: contig of 1151 bp in length
* 1152 1251: gap of unknown length
* 1252 2275: contig of 1024 bp in length
* 2276 2375: gap of unknown length
* 2376 4498: contig of 2123 bp in length
* 4499 4599: gap of unknown length
* 4599 6353: contig of 1755 bp in length
* 6354 6453: gap of unknown length
* 6454 8079: contig of 1626 bp in length
* 8080 8179: gap of unknown length
* 8180 30281: contig of 22102 bp in length
* 30282 30381: gap of unknown length
* 30382 65925: contig of 35544 bp in length
* 65926 66025: gap of unknown length
* 66026 107670: contig of 41645 bp in length
* 107671 107770: gap of unknown length
* 107771 167012: contig of 59242 bp in length.
Location/Qualifiers
1. 167012
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CITC-308K20"
/clone.lib="Caltech human BAC library C"
BASE COUNT 43272 a 39820 c 39446 g 43467 t 1007 others
ORIGIN

Query Match 12.4%; Score 158; DB 2; Length 167012;
Best Local Similarity 99.5%; Pred. No. 1.5e-69;
Matches 208; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 360 TCTGGAATAGTGGTCCATCCCTGAGACTGGGAGACCCAGGCAATGATTAACATA 419
DB 74627 TCTGGAATAGTGGTCCATCCCTGAGACTGGGAGACCCAGGCAATGATTAACATA 74686
OY 420 CTGGCTGTTGTGTGACCAATGATGTTGTGTAATTTGGAAGCAAGTATGTC 479
DB 74687 CTGGCTGTTGTGTGACCAATGATGTTGTGTAATTTGGAAGCAAGTATGTC 74746
OY 480 CAAGAGTGAATTAATTTGGCTGAAAGTTCATAGCATATTGGCTATGATTGACAGAGGG 539

Db 74747 CAAGAGTGAATTAATTTGGCTGAAAGTTCATAGCATATTGGCTATGATTGACAGAGGG 74806
OY 540 AAACCGACTGGAAGTCATTGCCATTAAAT 568
Db 74807 AAACCGACTGGAAGTCATTGCCATTAAAT 74835

RESULT 17
AC021159
LOCUS
DEFINITION
AC021159
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 213025)
Material: R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 213025)
Waterson, R.H.
Direct Submission
Submitted (14-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 19, 2000 this sequence version replaced gi:8570288.

COMMENT

NOTE: This is a 'working draft' sequence. It currently
* consists of 41 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1384: contig of 1384 bp in length
* 1385 1484: gap of unknown length
* 1485 3382: contig of 1898 bp in length
* 3383 3482: gap of unknown length
* 3483 5013: contig of 1531 bp in length
* 5014 5113: gap of unknown length
* 5114 6322: contig of 1209 bp in length
* 6323 6422: gap of unknown length
* 6423 7736: contig of 1314 bp in length
* 7737 7836: gap of unknown length
* 7837 9198: contig of 1362 bp in length
* 9199 9298: gap of unknown length
* 9299 11720: contig of 2422 bp in length
* 11721 11820: gap of unknown length
* 11821 13498: contig of 1678 bp in length
* 13499 13598: gap of unknown length
* 13599 15122: contig of 1524 bp in length
* 15123 15222: gap of unknown length
* 15223 17332: contig of 2110 bp in length
* 15224 17432: gap of unknown length
* 17433 19823: contig of 2391 bp in length
* 19824 19923: gap of unknown length
* 19924 21811: contig of 1888 bp in length
* 21812 21912: gap of unknown length
* 21913 24564: contig of 2653 bp in length
* 24565 24664: gap of unknown length
* 24665 27708: contig of 3044 bp in length
* 27709 27808: gap of unknown length
* 27809 30176: contig of 2368 bp in length
* 30177 30277: gap of unknown length
* 30278 33515: contig of 3239 bp in length
* 33516 33515: gap of unknown length

* 33616 37205: contig of 3590 bp in length
* 37206 37305: gap of unknown length
* 37306 39846: contig of 2541 bp in length
* 39847 39946: gap of unknown length
* 39947 42384: contig of 2438 bp in length
* 42385 42484: gap of unknown length
* 42485 46260: contig of 3776 bp in length
* 46261 46360: gap of unknown length
* 46361 49107: contig of 2747 bp in length
* 49108 49207: gap of unknown length
* 49208 54501: contig of 5294 bp in length
* 54502 54601: gap of unknown length
* 54602 59315: contig of 4714 bp in length
* 59316 59415: gap of unknown length
* 59416 65627: contig of 6212 bp in length
* 65628 65727: gap of unknown length
* 65728 72485: contig of 6758 bp in length
* 72486 72585: gap of unknown length
* 72586 77910: contig of 5325 bp in length
* 77911 78011: gap of unknown length
* 78011 84348: contig of 6338 bp in length
* 84349 84448: gap of unknown length
* 84449 90437: contig of 5989 bp in length
* 90438 90537: gap of unknown length
* 90538 97928: contig of 7391 bp in length
* 97929 98028: gap of unknown length
* 98029 105065: contig of 7037 bp in length
* 105066 105165: gap of unknown length
* 105166 111323: contig of 6158 bp in length
* 111324 111423: gap of unknown length
* 111424 117734: contig of 6311 bp in length
* 117735 117834: gap of unknown length
* 117835 125427: contig of 7593 bp in length
* 125428 125527: gap of unknown length
* 125528 134991: contig of 9464 bp in length
* 134992 135091: gap of unknown length
* 135092 143447: contig of 8356 bp in length
* 143448 143547: gap of unknown length
* 143548 156094: contig of 12547 bp in length
* 156095 156194: gap of unknown length
* 156195 157939: contig of 1745 bp in length
* 157940 158039: gap of unknown length
* 158040 175011: contig of 16972 bp in length
* 175012 175111: gap of unknown length
* 175112 187724: contig of 12613 bp in length
* 187725 187824: gap of unknown length
* 208210 208310: contig of 20386 bp in length
* 208311 213025: contig of 4715 bp in length.
* Location/Qualifiers
1. 213025
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-536N17"
1. 1384
misc_feature
/note="assembly_name:Contig11"
1485. 3382
misc_feature
/note="assembly_name:Contig14"
3483. 5013
misc_feature
/note="assembly_name:Contig15"
5114. 6322
misc_feature
/note="assembly_name:Contig16"
6423. 7736
misc_feature
/note="assembly_name:Contig18"
7837. 9198
misc_feature
/note="assembly_name:Contig24"
9299. 11720
misc_feature
/note="assembly_name:Contig25"
11821. 13498
misc_feature
/note="assembly_name:Contig26"
13599. 15122
misc_feature
/note="assembly_name:Contig27"

misc_feature 15223. 17332
/note="assembly_name:Contig28"
misc_feature 17433. 19823
/note="assembly_name:Contig30"
19924. 21811
misc_feature
/note="assembly_name:Contig31"
21912. 24564
misc_feature
/note="assembly_name:Contig32"
24665. 27708
misc_feature
/note="assembly_name:Contig33"
27809. 30176
misc_feature
/note="assembly_name:Contig34"
30277. 33515
misc_feature
/note="assembly_name:Contig36"
33616. 37205
misc_feature
/note="assembly_name:Contig37"
37306. 39846
misc_feature
/note="assembly_name:Contig38"
39947. 42384
misc_feature
/note="assembly_name:Contig39"
42485. 46260
misc_feature
/note="assembly_name:Contig40"
46361. 49107
misc_feature
/note="assembly_name:Contig41"
clone_end:SP6
vector_side:left"
49208. 54501
misc_feature
/note="assembly_name:Contig42"
54602. 59315
misc_feature
/note="assembly_name:Contig43"
59416. 65627
misc_feature
/note="assembly_name:Contig44"
65728. 72485
misc_feature
/note="assembly_name:Contig45"
72586. 77910
misc_feature
/note="assembly_name:Contig46"
78011. 84348
misc_feature
/note="assembly_name:Contig47"
84449. 90437
misc_feature
/note="assembly_name:Contig48"
90538. 97928
misc_feature
/note="assembly_name:Contig49"
98029. 105065
misc_feature
/note="assembly_name:Contig50"
105166. 111323
misc_feature
/note="assembly_name:Contig51"
111424. 117734
misc_feature
/note="assembly_name:Contig53"
117835. 125427
misc_feature
/note="assembly_name:Contig54"
125528. 134991
misc_feature
/note="assembly_name:Contig55"
135092. 143447
misc_feature
/note="assembly_name:Contig56"
143548. 156094
misc_feature
/note="assembly_name:Contig57"
156195. 157939
misc_feature
/note="assembly_name:Contig58"
158040. 175011
misc_feature
/note="assembly_name:Contig59"
175112. 187724
misc_feature
/note="assembly_name:Contig60"
187825. 208210
misc_feature
/note="assembly_name:Contig61"

Query Match 12.4% Score 158; DB 2; length 213025;
Best Local Similarity 99.5%; Pred. No. 1,5e-69;
Matches 208; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

0Y 360 TCTGGAAGTATGTCCTCCAGACTTGGGAAGCCAGGCGACATGATAAAGCTA 419
|||||
DB 75909 TCTGGAAGTATGTCCTCCAGACTTGGGAAGCCAGGCGACATGATAAAGCTA 75968

OY 420 CTGGCTGTGTGGTACATGACCCATGTGTAATTGGAAGCAAGTATGTG 479
|||||
DB 75969 CTGGCTGTGTGGTACATGACCCATGTGTAATTGGAAGCAAGTATGTG 76028
OY 480 CAAAGAGTAATTAATTCGCGTGAAGTCTTAGCATATGCTTGAATGCAAGGG 539
|||||
DB 76029 CAAAGAGTAATTAATTCGCGTGAAGTCTTAGCATATGCTTGAATGCAAGGG 76088
OY 540 AAACGACTGGAAGTCAATTCATTAAT 568
|||||
DB 76089 AAACGACTGGAAGTCAATTCATTAAT 76117
RESULT 18
AC067966 205656 bp DNA HTG 10-MAY-2000
AC067966/c
LOCUS
DEFINITION Homo sapiens chromosome CHROM 10 clone RP11-379018, *** SEQUENCING
IN PROGRESS ***
ACCESSION AC067966 GI:7767730
VERSION AC067966.2
KEYWORDS HTG; HTGS; PHASE1.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 205656)
JOURNAL Genome Therapeutics Corporation Sequencing Center: Human Genome
REFERENCE Smith, D.R.
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (27-APR-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT On May 10, 2000 this sequence version replaced gi:7656685.
----- Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: <http://www.genomecorp.com/>
Contact: gtc-seqcenter@genomecorp.com
----- Project Information
Center project name: hg286
----- Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 126548 bases at least Q40
Consensus quality: 159958 bases at least Q30
Consensus quality: 170609 bases at least Q20
Insert size: 205656; sum-of-contigs
Quality coverage: 2.4x in Q20 bases; sum-of-contigs
NOTE: This is a 'working draft' sequence. It currently
* consists of 117 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
I 1382: contig of 1382 bp in length
* 1383: gap of unknown length
* 2437: contig of 1055 bp in length
* 2438: gap of unknown length
* 3932: contig of 1495 bp in length
* 3933: gap of unknown length
* 5395: contig of 1463 bp in length
* 5396: gap of unknown length
* 6431: contig of 1036 bp in length
* 6432: gap of unknown length
* 7698: contig of 1267 bp in length

* 7699: gap of unknown length
* 8821: contig of 1123 bp in length
* 8822: gap of unknown length
* 9919: contig of 1098 bp in length
* 9920: gap of unknown length
* 11065: contig of 1146 bp in length
* 11066: gap of unknown length
* 12769: contig of 1704 bp in length
* 12770: gap of unknown length
* 14061: contig of 1292 bp in length
* 14062: gap of unknown length
* 15230: contig of 1169 bp in length
* 15231: gap of unknown length
* 16237: contig of 1007 bp in length
* 16238: gap of unknown length
* 17348: contig of 1111 bp in length
* 17349: gap of unknown length
* 18409: contig of 1061 bp in length
* 18410: gap of unknown length
* 19857: contig of 1448 bp in length
* 19858: gap of unknown length
* 21289: contig of 1432 bp in length
* 21290: gap of unknown length
* 22433: contig of 1144 bp in length
* 22434: gap of unknown length
* 23819: contig of 1386 bp in length
* 23820: gap of unknown length
* 25788: contig of 1969 bp in length
* 25789: gap of unknown length
* 26867: contig of 1079 bp in length
* 26868: gap of unknown length
* 28331: contig of 1464 bp in length
* 28332: gap of unknown length
* 29340: contig of 1009 bp in length
* 29341: gap of unknown length
* 30459: contig of 1119 bp in length
* 30460: gap of unknown length
* 31600: contig of 1141 bp in length
* 31601: gap of unknown length
* 32790: contig of 1190 bp in length
* 32791: gap of unknown length
* 33825: contig of 1035 bp in length
* 33826: gap of unknown length
* 35833: contig of 2008 bp in length
* 35834: gap of unknown length
* 37115: contig of 1282 bp in length
* 37116: gap of unknown length
* 38775: contig of 1660 bp in length
* 38776: gap of unknown length
* 40065: contig of 1290 bp in length
* 40066: gap of unknown length
* 41910: contig of 1845 bp in length
* 41911: gap of unknown length
* 44094: contig of 2184 bp in length
* 44095: gap of unknown length
* 45463: contig of 1369 bp in length
* 45464: gap of unknown length
* 46887: contig of 1424 bp in length
* 46888: gap of unknown length
* 48429: contig of 1542 bp in length
* 48430: gap of unknown length
* 49705: contig of 1276 bp in length
* 49706: gap of unknown length
* 51649: contig of 1944 bp in length
* 51650: gap of unknown length
* 52985: contig of 1336 bp in length
* 52986: gap of unknown length
* 54144: contig of 1159 bp in length
* 54145: gap of unknown length
* 55442: contig of 1298 bp in length
* 55443: gap of unknown length
* 57106: contig of 1664 bp in length
* 57107: gap of unknown length

Query Match 10.28; Score 130; DB 2; Length 205656;
Best Local Similarity 100.0%; Pred. No. 3; 2e-55;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match	Score	DB	Length	205656
58300	10.28	2	1194	bp in length
58301	10.28	2	1080	bp in length
59381	10.28	2	1167	bp in length
60548	10.28	2	1847	bp in length
62395	10.28	2	1039	bp in length
63434	10.28	2	1557	bp in length
64991	10.28	2	1416	bp in length
66407	10.28	2	1444	bp in length
67851	10.28	2	1401	bp in length
69252	10.28	2	1053	bp in length
70305	10.28	2	2050	bp in length
72355	10.28	2	1740	bp in length
74095	10.28	2	1839	bp in length
75934	10.28	2	1549	bp in length
77483	10.28	2	1319	bp in length
78802	10.28	2	1408	bp in length
80210	10.28	2	1529	bp in length
81739	10.28	2	1931	bp in length
83670	10.28	2	1206	bp in length
84876	10.28	2	1811	bp in length
86687	10.28	2	1031	bp in length
87718	10.28	2	1235	bp in length
88953	10.28	2	1955	bp in length
90908	10.28	2	1437	bp in length
92345	10.28	2	1496	bp in length
93841	10.28	2	1163	bp in length
95004	10.28	2	2204	bp in length
97208	10.28	2	1900	bp in length
99108	10.28	2	1831	bp in length
100939	10.28	2	2172	bp in length
103111	10.28	2	2092	bp in length
105203	10.28	2	1556	bp in length
106759	10.28	2	1344	bp in length
108103	10.28	2	1256	bp in length
109359	10.28	2	1525	bp in length

599 GATATCATGATGTCAAAGCGCTGAACCTGCTACTAGAGCTACTGCTGTTT 658
|||||
20249 GATATCATGATGTCAAAGCGCTGAACCTGCTACTAGAGCTACTGCTGTTT 20190
|||||
659 AGAAGTATAGGTTCTGATGATGAACACAGAAATGATGCTTAAATGCAAAATT 718
|||||
20189 AGAAGTATAGGTTCTGATGATGAACACAGAAATGATGCTTAAATGCAAAATT 20130
|||||
719 AAAGTATAGG 728
|||||
20129 AAAGTATAGG 20120
|||||
RESULT 19
AC067746
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT

AC067746 195126 bp DNA HTG 16-MAY-2001
Homo sapiens chromosome 10 clone RP11-408L20, WORKING DRAFT
SEQUENCE, 11 unordered pieces.
AC067746
AC067746.5 GI:14091814
HTG: HTGS_PHASE1: HTGS_DRAFT; HTGS_FULLTOP.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 195126)
Smith, D.R.
Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
Unpublished
2 (bases 1 to 195126)
Smith, D.R.
Direct Submission
Submitted (27-APR-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
On May 16, 2001 this sequence version replaced gi:8389438.
Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
Project Information
Center project name: hg288
Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 188219 bases at least Q40
Consensus quality: 189837 bases at least Q30
Consensus quality: 190569 bases at least Q20
Insert size: 194225; sum-of-contigs
Quality coverage: 6.7x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1102: contig of 1102 bp in length
1103 1202: gap of unknown length
1203 2394: contig of 1192 bp in length
2395 2494: gap of unknown length
2495 3744: contig of 1250 bp in length
3745 3844: gap of unknown length
3845 5030: contig of 1186 bp in length
5031 5130: gap of unknown length
5131 15992: contig of 10862 bp in length
15993 16093: gap of unknown length
16093 31383: contig of 15291 bp in length

FEATURES
source
1. 195126
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone_lib="RC1-11"
1. 1102
/note="assembly_name:Contig16"
1203. 2394
/note="assembly_name:Contig18"
2495. 3744
/note="assembly_name:Contig19"
3845. 5030
/note="assembly_name:Contig21"
5131. 15992
/note="assembly_name:Contig31"
16093. 31383
/note="assembly_name:Contig32"
31484. 47125
/note="assembly_name:Contig33"
clone_end:17"
47226. 69920
/note="assembly_name:Contig34"
70021. 92732
/note="assembly_name:Contig35"
clone_end:SP6"
92833. 118005
/note="assembly_name:Contig36"
118106. 195126
/note="assembly_name:Contig37"
BASE COUNT 52995 a 46055 c 45277 g 49797 t 1002 others
ORIGIN
Query Match 7.2%: Score 92; DB 2; Length 195126;
Best Local Similarity 100.0%; Pred. No. 8.5e-36;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 384 AGACTGGGAAGACCCAGGCGACATGATTAACATGCTGTGTGTGACATGACC 443
DB 2533 AGACTGGGAAGACCCAGGCGACATGATTAACATGCTGTGTGTGACATGACC 2592
OY 444 CAATTGATGTGTGTAATGGAAGCAAGSTA 475
DB 2593 CAATTGATGTGTGTAATGGAAGCAAGSTA 2624
RESULT 20
AC067749/c 215049 bp DNA HTG 26-MAY-2001
LOCUS AC067749.4 GI:14209706
DEFINITION Homo sapiens chromosome 10 clone RP11-367H5, WORKING DRAFT
SEQUENCE 33 unordered pieces.
ACCESSION AC067749
VERSION AC067749.4 GI:14209706
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 215049)
AUTHORS Smith,D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome

Sequence Data
Unpublished
2 (bases 1 to 215049)
REFERENCE
AUTHORS
TITLE
JOURNAL
DIRECT SUBMISSION
Submitted (27-APR-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
On May 26, 2001 this sequence version replaced gi:8954013.
COMMENT
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
Project Information
Center project name: hg285
Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 194485 bases at least Q40
Consensus quality: 199162 bases at least Q30
Consensus quality: 201282 bases at least Q20
Insert size: 211948; sum-of-coverage
Quality coverage: 6.8x in Q20 bases; sum-of-coverage
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1163: contig of 1163 bp in length
1 1164: gap of unknown length
1 1263: contig of 1013 bp in length
1 1264: gap of unknown length
1 2277: contig of 1079 bp in length
1 2377: gap of unknown length
1 3455: contig of 1032 bp in length
1 3555: gap of unknown length
1 3556: contig of 1032 bp in length
1 4587: gap of unknown length
1 4588: contig of 1008 bp in length
1 4688: gap of unknown length
1 5595: contig of 1175 bp in length
1 5795: gap of unknown length
1 5796: contig of 1175 bp in length
1 6970: gap of unknown length
1 6971: contig of 1087 bp in length
1 7071: gap of unknown length
1 8157: contig of 1043 bp in length
1 8257: gap of unknown length
1 8258: contig of 1043 bp in length
1 8259: gap of unknown length
1 9301: contig of 1041 bp in length
1 9401: gap of unknown length
1 10441: contig of 1041 bp in length
1 10442: gap of unknown length
1 10541: contig of 1122 bp in length
1 1163: contig of 1122 bp in length
1 1164: gap of unknown length
1 1165: contig of 1163 bp in length
1 11764: gap of unknown length
1 12926: contig of 1163 bp in length
1 12927: gap of unknown length
1 13027: contig of 1142 bp in length
1 14169: gap of unknown length
1 14269: contig of 1306 bp in length
1 14270: gap of unknown length
1 15574: contig of 1306 bp in length
1 15575: gap of unknown length
1 15576: contig of 1364 bp in length
1 15675: gap of unknown length
1 16738: contig of 1263 bp in length
1 17039: gap of unknown length
1 17139: contig of 1263 bp in length
1 18401: gap of unknown length
1 18501: contig of 1287 bp in length
1 18502: gap of unknown length
1 19788: contig of 1287 bp in length
1 19789: gap of unknown length
1 19889: contig of 157 bp in length
1 19890: gap of unknown length
1 21246: contig of 1722 bp in length
1 21346: gap of unknown length
1 23067: contig of 1722 bp in length
1 23167: gap of unknown length
1 25167: contig of 2000 bp in length
1 25168: gap of unknown length
1 25169: contig of 2399 bp in length
1 25267: gap of unknown length
1 28266: contig of 2399 bp in length
1 28267: gap of unknown length
1 28366: gap of unknown length


```
* 28367 30737: contig of 2371 bp in length
* 30738 30837: gap of unknown length
* 30838 35830: contig of 4993 bp in length
* 35831 35930: gap of unknown length
* 35931 43873: contig of 7943 bp in length
* 43874 43973: gap of unknown length
* 43974 54313: contig of 10340 bp in length
* 54314 62524: gap of unknown length
* 62525 62625: gap of unknown length
* 62625 75784: contig of 13160 bp in length
* 75785 75885: gap of unknown length
* 75885 92445: contig of 16561 bp in length
* 92446 92546: gap of unknown length
* 92546 109803: contig of 17258 bp in length
* 109804 109903: gap of unknown length
* 109904 128494: contig of 18591 bp in length
* 128495 128594: gap of unknown length
* 128595 146740: contig of 18145 bp in length
* 146740 146839: gap of unknown length
* 146840 167230: contig of 20391 bp in length
* 167231 167331: gap of unknown length
* 167331 187365: contig of 20035 bp in length
* 187366 187466: gap of unknown length
* 187466 215049: contig of 27584 bp in length.
*
Location/Qualifiers
1. 215049
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone_id="RP11-367H5"
/clone_id="RP11-11"
1. 1163
/note="assembly_name:Contig4"
1264. 2276
/note="assembly_name:Contig4"
misc_feature
2377. 3435
/note="assembly_name:Contig7"
misc_feature
3556. 4587
/note="assembly_name:Contig12"
misc_feature
4688. 5695
/note="assembly_name:Contig15"
misc_feature
5796. 6970
/note="assembly_name:Contig55"
misc_feature
7071. 8157
/note="assembly_name:Contig60"
misc_feature
8258. 9300
/note="assembly_name:Contig62"
misc_feature
9401. 10441
/note="assembly_name:Contig63"
misc_feature
10542. 11663
/note="assembly_name:Contig69"
misc_feature
11764. 12926
/note="assembly_name:Contig73"
misc_feature
13027. 14168
/note="assembly_name:Contig75"
misc_feature
14269. 15574
/note="assembly_name:Contig112"
misc_feature
15675. 17038
/note="assembly_name:Contig113"
misc_feature
17139. 18401
/note="assembly_name:Contig116"
misc_feature
18502. 19788
/note="assembly_name:Contig118"
misc_feature
19889. 21245
/note="assembly_name:Contig121"
misc_feature
21346. 23067
/note="assembly_name:Contig124"
misc_feature
23168. 25167
/note="assembly_name:Contig127"
misc_feature
25268. 28266
/note="assembly_name:Contig134"
misc_feature
25268. 28266
/note="assembly_name:Contig135"
misc_feature
```

```
misc_feature 28367. 30737
/note="assembly_name:Contig136"
misc_feature 30838. 35830
/note="assembly_name:Contig137"
misc_feature 35931. 43873
/note="assembly_name:Contig138"
misc_feature 43974. 54313
/note="assembly_name:Contig139"
misc_feature 54414. 62524
/note="assembly_name:Contig140"
misc_feature 62625. 75784
/note="assembly_name:Contig141"
misc_feature 75885. 92445
/note="assembly_name:Contig142"
misc_feature 92546. 109803
/note="assembly_name:Contig143"
misc_feature 109904. 128494
/note="assembly_name:Contig144"
misc_feature 128595. 146739
/note="assembly_name:Contig145"
misc_feature 146840. 167230
/note="assembly_name:Contig146"
misc_feature 167331. 187365
/note="assembly_name:Contig147"
misc_feature 187466. 215049
/note="assembly_name:Contig148"
clone_end:"7"

BASE COUNT 57394 a 47838 c 47126 g 59466 t 3225 others
ORIGIN
```

```
Query Match 7.2%; Score 92; DB 2; Length 215049;
Best Local Similarity 100.0%; Pred. No. 8.56-36;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 599 GATATCATGATGCTCAACGCTGAACCTGCTACTTGAAGCTACTGTGACACTGTTT 658
|||||
DB 47543 GATATCATGATGCTCAACGCTGAACCTGCTACTTGAAGCTACTGTGACACTGTTT 47484
|||||
QY 659 ACAAGGTATTAAGTTCCTGATGGAACCAAGAAATGATTTGGCTTAATGACAGATT 718
|||||
DB 47483 ACAAGGTATTAAGTTCCTGATGGAACCAAGAAATGATTTGGCTTAATGACAGATT 47424
|||||
QY 719 AAAGATTAAGG 728
|||||
DB 47423 AAAGATTAAGG 47414
```

```
RESULT 21
AC011012 155645 bp DNA HTG 03-NOV-2000
LOCUS Homo sapiens clone RP11-8023, WORKING DRAFT SEQUENCE, 7 unordered
DEFINITION pieces.
ACCESSION AC011012
VERSION AC011012.4 GI:11079562
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 155645)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Baines,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferrelia,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gaidyna,S., Grant,G., Hagos,B., Heathford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
```


* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 849: contig of 849 bp in length
850 949: gap of 100 bp
950 1823: contig of 874 bp in length
1824 1923: gap of 100 bp
1924 2761: contig of 838 bp in length
2762 2861: gap of 100 bp
2862 3738: contig of 877 bp in length
3739 3838: gap of 100 bp
3839 4684: contig of 846 bp in length
4685 4784: gap of 100 bp
4785 5604: contig of 820 bp in length
5605 5704: gap of 100 bp
5705 6518: contig of 814 bp in length
6519 6618: gap of 100 bp
6619 7445: contig of 827 bp in length
7446 7545: gap of 100 bp
7546 8391: contig of 846 bp in length
8392 8491: gap of 100 bp
8492 9312: contig of 821 bp in length
9313 9412: gap of 100 bp
9413 10228: contig of 816 bp in length
10229 10328: gap of 100 bp
10329 11170: contig of 842 bp in length
11171 11270: gap of 100 bp
11271 12097: contig of 827 bp in length
12098 12197: gap of 100 bp
12198 13031: contig of 834 bp in length
13032 13131: gap of 100 bp
13132 13940: contig of 809 bp in length
13941 14040: gap of 100 bp
14041 14887: contig of 847 bp in length
14888 14987: gap of 100 bp
14988 15823: contig of 836 bp in length
15824 15923: gap of 100 bp
15924 16748: contig of 825 bp in length
16749 16848: gap of 100 bp
16849 17692: contig of 844 bp in length
17693 17792: gap of 100 bp
17793 18636: contig of 844 bp in length
18637 18736: gap of 100 bp
18737 19596: contig of 860 bp in length
19597 19696: gap of 100 bp
19697 20534: contig of 838 bp in length
20535 20634: gap of 100 bp
20635 21446: contig of 812 bp in length
21447 21546: gap of 100 bp
21547 22391: contig of 845 bp in length
22392 22491: gap of 100 bp
22492 23319: contig of 828 bp in length
23320 23419: gap of 100 bp
23420 24268: contig of 849 bp in length
24269 24368: gap of 100 bp
24369 25218: contig of 850 bp in length
25219 25318: gap of 100 bp
25319 26161: contig of 843 bp in length
26162 26261: gap of 100 bp
26262 27120: contig of 859 bp in length
27121 27220: gap of 100 bp
27221 28039: contig of 819 bp in length
28040 28139: gap of 100 bp
28140 28978: contig of 839 bp in length
28979 29078: gap of 100 bp
29079 29937: contig of 859 bp in length
29938 30037: gap of 100 bp
30038 30860: contig of 823 bp in length

30861 30960: gap of 100 bp
30961 31777: contig of 817 bp in length
31778 31877: gap of 100 bp
31878 32749: contig of 872 bp in length
32750 32849: gap of 100 bp
32850 33680: contig of 831 bp in length
33681 33780: gap of 100 bp
33781 34625: contig of 845 bp in length
34626 34725: gap of 100 bp
34726 35561: contig of 836 bp in length
35562 35661: gap of 100 bp
35662 36540: contig of 879 bp in length
36541 36640: gap of 100 bp
36641 37473: contig of 833 bp in length
37474 37573: gap of 100 bp
37574 38378: contig of 805 bp in length
38379 38478: gap of 100 bp
38479 39345: contig of 867 bp in length
39346 39445: gap of 100 bp
39446 40238: contig of 793 bp in length
40239 40338: gap of 100 bp
40339 41169: contig of 831 bp in length
41170 41269: gap of 100 bp
41270 42089: contig of 820 bp in length
42090 42189: gap of 100 bp
42190 43054: contig of 865 bp in length
43055 43154: gap of 100 bp
43155 44013: contig of 859 bp in length
44014 44113: gap of 100 bp
44114 44982: contig of 869 bp in length
44983 45082: gap of 100 bp
45083 45956: contig of 874 bp in length
45957 46056: gap of 100 bp
46057 46687: contig of 831 bp in length
46688 46987: gap of 100 bp
46988 47800: contig of 813 bp in length
47801 47900: gap of 100 bp
47901 48759: contig of 859 bp in length
48760 48859: gap of 100 bp
48860 49719: contig of 860 bp in length
49720 49819: gap of 100 bp
49820 50690: contig of 871 bp in length
50691 50790: gap of 100 bp
50791 51641: contig of 851 bp in length
51642 51741: gap of 100 bp
51742 52602: contig of 861 bp in length
52603 52702: gap of 100 bp
52703 53577: contig of 875 bp in length
53578 53677: gap of 100 bp
53678 54522: contig of 845 bp in length
54523 54622: gap of 100 bp
54623 55470: contig of 848 bp in length
55471 55570: gap of 100 bp
55571 56408: contig of 838 bp in length
56409 56508: gap of 100 bp
56509 57360: contig of 852 bp in length
57361 57460: gap of 100 bp
57461 58286: contig of 826 bp in length
58287 58386: gap of 100 bp
58387 59224: contig of 838 bp in length
59225 59324: gap of 100 bp
59325 60131: contig of 807 bp in length
60132 60231: gap of 100 bp
60232 61082: contig of 851 bp in length
61083 61182: gap of 100 bp
61183 62027: contig of 845 bp in length
62028 62127: gap of 100 bp
62128 62985: contig of 828 bp in length
62986 63085: gap of 100 bp
63086 63913: contig of 828 bp in length
63914 64013: gap of 100 bp
64014 64873: contig of 860 bp in length
64874 64973: gap of 100 bp

* 64974 65814: contig of 841 bp in length

Query Match 5.8%; Score 74; DB 2; Length 70575;

Best Local Similarity 100.0%; Pred. No. 1.4e-26;

Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 425 TGTGTGGGACACATGACCAATGATGCTGTGAAATGGAAGAGGTATGTGACAGA 484

DB 33900 TGTGTGGGACACATGACCAATGATGCTGTGAAATGGAAGAGGTATGTGACAGA 33959

OY 485 GGTGAATTAATTGG 498

DB 33960 GGTGAATTAATTGG 33973

RESULT 23

BOVIOPPP

LOCUS BOVIOPPP 1266 bp mRNA MAM 27-APR-1993

DEFINITION

BOVINE INORGANIC PYROPHOSPHATASE mRNA sequence.

ACCESSION

M95283

VERSION

M95283.1 GI:163228

KEYWORDS

Inorganic pyrophosphatase.

SOURCE

Bos taurus retina cDNA to mRNA.

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

1 (bases 1 to 1266)

REFERENCE

Yang, Z. and Wensel, T. G.

Molecular cloning and functional expression of cDNA encoding a

mammalian inorganic pyrophosphatase

J. Biol. Chem. 267, 24641-24647 (1992)

JOURNAL

FEATURES

Location/Qualifiers

Source

1..1266

/organism="Bos taurus"

/db_xref="taxon:9913"

/tissue_type="retina"

BASE COUNT

396 a 239 c 300 g 331 t

ORIGIN

Query Match 4.4%; Score 56; DB 4; Length 1266;

Best Local Similarity 100.0%; Pred. No. 2.6e-17;

Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 419 ACTGGCTGTGTGTGACCAATGACCAATGATGCTGTGAAATGGAAGAGGT 474

DB 399 ACTGGCTGTGTGTGACCAATGACCAATGATGCTGTGAAATGGAAGAGGT 454

RESULT 24

BC010468

LOCUS BC010468 1243 bp mRNA ROD 12-JUL-2001

DEFINITION

Mus musculus, RIKEN cDNA 2010317E03 gene, clone MGC:6716

ACCESSION

BC010468

VERSION

BC010468.1 GI:14714656

KEYWORDS

MGC.

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

1 (bases 1 to 1243)

REFERENCE

Strausberg, R.

Direct Submission

Submitted (10-JUL-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgaps-remail.nih.gov

REMARK

COMMENT

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: villalobos@bcm.tmc.edu

Villalon, D.K., Luna, R.A., Hale, S.M., Huylk, S., Lu, X., Garcia,

A.M., Hollway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,

Muzny, D.M., Gibbs, R.A.

Clone distribution: MCC clone distribution information can be found

through the I.M.A.G.E. Consortium/HLN at: <http://image.llnl.gov>

Series: IRAC Plate: 10 Row: f Column: 15

This clone was selected for full length sequencing because it

passed the following selection criteria: Genomescan gene

prediction, Similarity but not identity to protein.

Location/Qualifiers

1..1243

/organism="Mus musculus"

/db_xref="locusid:67895"

/db_xref="taxon:10090"

/clone="MGC:6716 IMAGE:3585780"

/tissue_type="Mammary tumor. Metallothionlen-TGF alpha

model. 10 month Old. Virgin mouse. Taken by biopsy."

/clone_id="NCI_CGAP_Mam1"

/lab_host="DH10B"

/note="Vector: PCMV-SPORT6"

42..911

/codon_start=1

/product="RIKEN cDNA 2010317E03 gene"

/protein_id="AAH10468.1"

/db_xref="GI:14714657"

/translation="MSGFSSSEERAAPEFLRYFLKNEGQVYSPHVDPIADKDF

HMVEVPRMSNAKMEIATKDPINIKODVYKIRYANLPYKGIWNGAIPQWE

DPGSDKHGCGCGNDPIDCEISKVCAGREIIRVGLIAMDIGETDVIAMIN

VDDPDANKDISVEYELAKGYLEATYDMPRRKRVKPGRENFAPNAEKKNDFAVD

IKSTHYWKALVTKTDGKISCMNTTVSESPKCPDPAKAIYDALPPCESACSL

PVDVDFWPHOOKN"

BASE COUNT

366 a 270 c 310 g 297 t

ORIGIN

Query Match 3.9%; Score 50; DB 10; Length 1243;

Best Local Similarity 100.0%; Pred. No. 3e-14;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 653 TGGTTAGAGGTATAGGTTCTGATGGAAGAACGAGAAATGAGTTGC 702

DB 606 TGGTTAGAGGTATAGGTTCTGATGGAAGAACGAGAAATGAGTTGC 655

RESULT 25

AC025103/C

LOCUS AC025103/C 70575 bp DNA HTG 13-JUL-2000

DEFINITION

Homo sapiens clone RP11-25A15, LOW-PASS SEQUENCE SAMPLING.

ACCESSION

AC025103

VERSION

AC025103.1 GI:7158913

KEYWORDS

HTG; HTGS_PHASE0.

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 70575)

REFERENCE

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Unpublished

2 (bases 1 to 70575)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Balwin, J., Barina, N., Bastien, V., Bida, F.,

Boguslavsky, L., Bouknight, B., Brown, A., Buxton, G.,

Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,

Conrad, T., Delella, D., Drenth, J., Drenth, J., Drenth, J.,

Drenth, J., Drenth, J., Drenth, J., Drenth, J., Drenth, J.,

Drenth, J., Drenth, J., Drenth, J., Drenth, J., Drenth, J.,

TITLE
JOURNAL
COMMENT

Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, M., Gage, D.,
Gallagan, J., Gardyna, S., Glend, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hages, B., Hearford, A., Horton, L.,
Howland, J.C., Illey, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehouck, J.,
Levine, R., Lien, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McPheters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mienda, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisanal, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rotman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:

Slit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L4527

Center clone name: 25_A_15

NOTE: This record contains 75 individual
sequencing reads that have not been assembled into
contigs. Runs of N are used to separate the reads
and the order in which they appear is completely
arbitrary. Low-pass sequence sampling is useful for
identifying clones that may be gene-rich and allows
overlap relationships among clones to be deduced.
However, it should not be assumed that this clone
will be sequenced to completion. In the event that
the record is updated, the accession number will
be preserved.

1 849: contig of 849 bp in length
850 949: gap of 100 bp
950 1823: contig of 874 bp in length
1824 1923: gap of 100 bp
1924 2761: contig of 838 bp in length
2762 2861: gap of 100 bp
2862 3738: contig of 877 bp in length
3739 3838: gap of 100 bp
3839 4684: contig of 846 bp in length
4685 4784: gap of 100 bp
4785 5604: contig of 820 bp in length
5605 5704: gap of 100 bp
5705 6518: contig of 814 bp in length
6519 6618: gap of 100 bp
6619 7445: contig of 827 bp in length
7446 7545: gap of 100 bp
7546 8391: contig of 846 bp in length
8392 8491: gap of 100 bp
8492 9312: contig of 821 bp in length
9313 9412: gap of 100 bp
9413 10228: contig of 816 bp in length
10229 10328: gap of 100 bp
10329 11170: contig of 842 bp in length
11171 11270: gap of 100 bp
11271 12097: contig of 827 bp in length
12098 12197: gap of 100 bp
12198 13031: contig of 834 bp in length
13032 13131: gap of 100 bp
13132 13940: contig of 809 bp in length
13941 14040: gap of 100 bp

14041 14887: contig of 847 bp in length
14888 14987: gap of 100 bp
14988 15823: contig of 836 bp in length
15824 15923: gap of 100 bp
15924 16748: contig of 825 bp in length
16749 16848: gap of 100 bp
16849 17692: contig of 844 bp in length
17693 17792: gap of 100 bp
17793 18635: contig of 844 bp in length
18637 18736: gap of 100 bp
18737 18596: contig of 860 bp in length
18597 19696: gap of 100 bp
19697 20534: contig of 838 bp in length
20535 20634: gap of 100 bp
20635 21446: contig of 812 bp in length
21447 21546: gap of 100 bp
21547 22391: contig of 845 bp in length
22392 22491: gap of 100 bp
22492 23319: contig of 828 bp in length
23320 23419: gap of 100 bp
23420 24268: contig of 849 bp in length
24269 24368: gap of 100 bp
24369 25218: contig of 830 bp in length
25219 25318: gap of 100 bp
25319 26161: contig of 843 bp in length
26162 26261: gap of 100 bp
26262 27120: contig of 859 bp in length
27121 27220: gap of 100 bp
27221 28039: contig of 819 bp in length
28040 28139: gap of 100 bp
28140 28978: contig of 839 bp in length
28979 29078: gap of 100 bp
29079 29937: contig of 859 bp in length
29938 30037: gap of 100 bp
30038 30860: contig of 823 bp in length
30861 30960: gap of 100 bp
30961 31777: contig of 817 bp in length
31778 31877: gap of 100 bp
31878 32748: contig of 872 bp in length
32750 32845: gap of 100 bp
32850 33680: contig of 831 bp in length
33681 33780: gap of 100 bp
33781 34625: contig of 845 bp in length
34626 34725: gap of 100 bp
34726 35561: contig of 836 bp in length
35562 35661: gap of 100 bp
35662 36540: contig of 879 bp in length
36541 36640: gap of 100 bp
36641 37473: contig of 833 bp in length
37474 37573: gap of 100 bp
37574 38378: contig of 865 bp in length
38379 38478: gap of 100 bp
38479 39345: contig of 867 bp in length
39346 39445: gap of 100 bp
39446 40238: contig of 793 bp in length
40239 40338: gap of 100 bp
40339 41169: contig of 831 bp in length
41170 41269: gap of 100 bp
41270 42089: contig of 820 bp in length
42090 42189: gap of 100 bp
42190 43054: contig of 865 bp in length
43055 43154: gap of 100 bp
43155 44013: contig of 859 bp in length
44014 44113: gap of 100 bp
44114 44982: contig of 869 bp in length
44983 45082: gap of 100 bp
45083 45956: contig of 874 bp in length
45957 46056: gap of 100 bp
46057 46887: contig of 831 bp in length
46888 46987: gap of 100 bp
46988 47800: contig of 813 bp in length
47801 47900: gap of 100 bp
47901 48759: contig of 859 bp in length

```

* 48760 48859: gap of 100 bp
* 48860 49719: contig of 860 bp in length
* 49720 49819: gap of 100 bp
* 49820 50690: contig of 871 bp in length
* 50691 50790: gap of 100 bp
* 50791 51641: contig of 851 bp in length
* 51642 51741: gap of 100 bp
* 51742 52602: contig of 861 bp in length
* 52603 52702: gap of 100 bp
* 52703 53577: contig of 875 bp in length
* 53578 53677: gap of 100 bp
* 53678 54522: contig of 845 bp in length
* 54523 54622: gap of 100 bp
* 54623 55470: contig of 848 bp in length
* 55471 55570: gap of 100 bp
* 55571 56408: contig of 838 bp in length
* 56409 56508: gap of 100 bp
* 56509 57360: contig of 852 bp in length
* 57361 57460: gap of 100 bp
* 57461 58286: contig of 826 bp in length
* 58287 58386: gap of 100 bp
* 58387 59224: contig of 838 bp in length
* 59225 59324: gap of 100 bp
* 59325 60131: contig of 807 bp in length
* 60132 60231: gap of 100 bp
* 60232 61082: contig of 851 bp in length
* 61083 61182: gap of 100 bp
* 61183 62027: contig of 845 bp in length
* 62028 62127: gap of 100 bp
* 62128 62985: contig of 858 bp in length
* 62986 63085: gap of 100 bp
* 63086 63913: contig of 828 bp in length
* 63914 64013: gap of 100 bp
* 64014 64873: contig of 860 bp in length
* 64874 64973: gap of 100 bp
* 64974 65814: contig of 841 bp in length

```

Query Match 3.5%; Score 44; DB 2; Length 70575;
 Best Local Similarity 100.0%; Pred. No. 3.1e-11;
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 GAAACCTGCTAGTTCGCAATTTGTCGCCGATTAAGGATAT 358
 Db 48087 GAAACCTGCTAGTTCGCAATTTGTCGCCGATTAAGGATAT 48044

RESULT 26
 AL513487/c AL513487 144177 bp DNA PRI 09-MAR-2001
 LOCUS Human DNA sequence from clone RP13-63115 on chromosome Xq24-26.2,
 DEFINITION complete sequence.
 ACCESSION AL513487
 VERSION AL513487.3 GI:13274904
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 144177)
 AUTHORS Heath, P.
 TITLE Direct Submission
 JOURNAL Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequests@sanger.ac.uk

COMMENT
 On Mar 12, 2001 this sequence version replaced gi:12831924.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate

FEATURES

source

chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPeP; Information on the WormPeP database can be found at <http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/REG/chrX> RP13-63115 is from the library RP13-13.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTER: pBAC3.6

This sequence is the entire insert of clone RP13-63115.

Location/Qualifiers

1..144177

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="q24-26.2"
 /clone="RP13-63115"
 /clone_id="RP13-13.1"
 12..660
 /note="L1M2 repeat: matches 1189..1850 of consensus"
 703..1121
 /note="L1P repeat: matches 1681..2103 of consensus"
 1193..1343
 /note="L1P15-16 repeat: matches 1023..1176 of consensus"
 1350..1822
 /note="L1P15-16 repeat: matches 425..886 of consensus"
 2031..2304
 /note="L1P15-16 repeat: matches 109..378 of consensus"
 2664..2875
 /note="L1P12 repeat: matches -1422..-1209 of consensus"
 2896..3326
 /note="L1M4c repeat: matches 1398..1849 of consensus"
 3328..3351
 /note="12 copies 2 mer tg 95% conserved"
 3353..3380
 /note="14 copies 2 mer ac 100% conserved"
 3422..3693
 /note="AluSc repeat: matches 1..273 of consensus"
 3694..4219
 /note="L1M4c repeat: matches 844..1412 of consensus"
 4233..4375
 /note="THE1B repeat: matches 3..159 of consensus"
 4374..4470
 /note="THE1B repeat: matches 49..159 of consensus"
 4471..5423
 /note="THE1B-INTERNAL repeat: matches 510..1571 of consensus"
 5424..5786
 /note="THE1B repeat: matches 2..364 of consensus"
 5865..5983
 /note="L1M4c repeat: matches 793..913 of consensus"
 6009..6319
 /note="AluSc repeat: matches 1..312 of consensus"
 6824..6882
 /note="L1M3c repeat: matches 11..67 of consensus"
 7277..10034
 /note="L1P15 repeat: matches 3280..6157 of consensus"
 10035..10343
 /note="Alu repeat: matches 1..302 of consensus"
 10344..12265
 /note="L1P15 repeat: matches 1412..3280 of consensus"
 14439..14804
 /note="THE1B repeat: matches 1..364 of consensus"
 14805..16355

```

/note="THE1B-INTERNAL repeat: matches 1. .1580 of
consensus"
repeat_region
16356. .16721
/note="THE1B repeat: matches 1. .364 of consensus"
repeat_region
17289. .17320
/note="16 copies 2 mer tt 84% conserved"
repeat_region
18611. .18725
/note="LIM4 repeat: matches 4363. .4484 of consensus"
repeat_region
20495. .20649
/note="LIP16 repeat: matches 4305. .4448 of consensus"
repeat_region
20702. .22403
/note="LIP16 repeat: matches 4444. .6157 of consensus"
repeat_region
22625. .22878
/note="MIR repeat: matches 13. .262 of consensus"
repeat_region
24431. .24916
/note="L2 repeat: matches 1778. .2276 of consensus"
repeat_region
25602. .26550
/note="L2 repeat: matches 1744. .2741 of consensus"
repeat_region
26745. .27153
/note="LIP repeat: matches 1450. .1858 of consensus"
repeat_region
27138. .28917
/note="LIP3 repeat: matches 4051. .5838 of consensus"
repeat_region
28918. .29237
/note="LIP3 repeat: matches 5826. .6146 of consensus"
repeat_region
29259. .33467
/note="LIP8 repeat: matches 1843. .6163 of consensus"
repeat_region
33618. .33724
/note="LIM4A repeat: matches 6194. .6298 of consensus"
repeat_region
34017. .34962
/note="LIP2 repeat: matches 5195. .6155 of consensus"
repeat_region
35433. .35654
/note="MIR repeat: matches 22. .249 of consensus"
repeat_region
35691. .35757
/note="LIM8 repeat: matches 6099. .6170 of consensus"
repeat_region
36190. .36444
/note="LIP2 repeat: matches 5892. .6146 of consensus"
repeat_region
37225. .37318
/note="47 copies 2 mer ta 68% conserved"
repeat_region
38864. .39056
/note="MIR repeat: matches 42. .252 of consensus"
repeat_region
39283. .39349
/note="L2 repeat: matches 2595. .2668 of consensus"
repeat_region
39591. .41301
/note="LIME2 repeat: matches 4395. .6137 of consensus"
repeat_region
41295. .42595
/note="LIP3 repeat: matches 4770. .6117 of consensus"
repeat_region
42611. .42634
/note="LIME2 repeat: matches 4374. .4396 of consensus"
repeat_region
42635. .43740
/note="MER1C repeat: matches 3. .1071 of consensus"
repeat_region
43741. .44993
/note="LIME2 repeat: matches 3112. .4374 of consensus"
repeat_region
45006. .45402
/note="LIP7 repeat: matches 5152. .5564 of consensus"
repeat_region
45403. .45973
/note="LIP7 repeat: matches 5552. .6141 of consensus"
repeat_region
46308. .46671
/note="LIP16 repeat: matches 5824. .6150 of consensus"
repeat_region
47439. .47743
/note="AluY repeat: matches 1. .304 of consensus"
repeat_region
49898. .50115
/note="MIR repeat: matches 33. .260 of consensus"
repeat_region
50777. .51235
/note="LIM9 repeat: matches 5837. .6306 of consensus"
repeat_region
52537. .52671
/note="LIM7 repeat: matches 6149. .6284 of consensus"
repeat_region
52998. .53083
/note="L2 repeat: matches 2295. .2380 of consensus"
repeat_region
53176. .53286
/note="L2 repeat: matches 2604. .2716 of consensus"
repeat_region
54010. .54049
/note="20 copies 2 mer ca 100% conserved"
repeat_region
54055. .54082

```

```

/note="14 copies 2 mer ag 100% conserved"
repeat_region
54126. .54204
/note="MAD1 repeat: matches 2. .80 of consensus"
repeat_region
54804. .54897
/note="MERSA repeat: matches 1. .90 of consensus"
repeat_region
55936. .56044
/note="MIR repeat: matches 35. .150 of consensus"
repeat_region
56467. .56656
/note="MERS repeat: matches 1. .209 of consensus"
repeat_region
57324. .57446
/note="LIMC repeat: matches 1927. .2055 of consensus"
repeat_region
58142. .58455
/note="AlusG repeat: matches 1. .309 of consensus"
repeat_region
58522. .58764
/note="MIR repeat: matches 12. .261 of consensus"
repeat_region
58904. .59119
/note="MIR repeat: matches 32. .258 of consensus"
repeat_region
59854. .60143
/note="145 copies 2 mer aa 55% conserved"
repeat_region
63213. .63935
/note="LIM8 repeat: matches 5534. .6285 of consensus"
repeat_region
64586. .65073
/note="M11J repeat: matches 7. .516 of consensus"
repeat_region
67119. .67162
/note="L2 repeat: matches 2645. .2692 of consensus"
repeat_region
67180. .67268
/note="LIMC4 repeat: matches 6608. .6696 of consensus"
repeat_region
67307. .67648
/note="LIMC4 repeat: matches 6271. .6611 of consensus"

```

```

Query Match 2.0%; Score 25; DB 9; Length 144177;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1237 ATATGTTACCAATGTTAGGAATA 1261
Db 64051 ATATGTTACCAATGTTAGGAATA 64027

```

```

RESULT 27
AF020802/c 61946 bp DNA PRI 05-MAY-2001
LOCUS Homo sapiens chromosome 21 clone cosmids q98a3 and c103c0352 map
DEFINITION 21q22.2, complete sequence.
ACCESSION AF020802
VERSION AF020802.2 GI:13957611
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE 1 (bases 1 to 61946)
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.-S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.-K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Wenzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M.B., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Kawasaki,K., Asakawa,S.,
Shintani,A., Sasaki,T., Nagamine,K., Mitsuyma,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G.,
Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Riesselmann,L., Dagand,E., Haef,T., Wehrmeyer,S.,
Reinhardt,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H.,
Boehm,R. and Yaspo,M.Laure.

```

```

TITLE The DNA sequence of human chromosome 21. The chromosome 21 mapping
and sequencing consortium
Nature 405 (6784), 311-319 (2000)

```

```

JOURNAL 2 (bases 1 to 61946)
MEDLINE Blechschmidt,K., Nordstiek,G., Drescher,B., Weber,J., Schattvov,R.,
REFERENCE Rosenthal,A., Yaspo,M.-L., Osoegawa,K. and Soeda,E.
AUTHORS Direct Submission
TITLE

```


JOURNAL Submitted (27-AUG-1997) Genome Analysis, Institute for Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
 REFERENCE 3 (bases 1 to 61946)
 AUTHORS Blechschmidt, K., Nordstieck, G., Drescher, B., Weber, J., Schattevoy, R., Rosenthal, A., Yaspo, M.-L., Osoegawa, K. and Soeda, E.
 TITLE Direct Submission
 JOURNAL Submitted (05-MAY-2001) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
 COMMENT On May 5, 2001 this sequence version replaced gi:2801436.
 FEATURES
 source
 1. 61946
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.2"
 /clone="cosmids g98A3 and c103C0352"
 BASE COUNT 19958 a 12828 c 12287 g 16873 t
 ORIGIN

Query Match 1.7%; Score 22; DB 9; Length 61946;
 Best Local Similarity 100.0%; Pred. No. 5.5;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 681 GAAACCAAGAAATGACTTGGC 702
 Db 4711 GAAACCAAGAAATGACTTGGC 4690

RESULT 28
 HS1042K10 184180 bp DNA PRI 12-DEC-1999
 LOCUS
 DEFINITION Human DNA sequence from clone RP5-1042K10 on chromosome 22q13.1-13.2. Contains the ADL gene for adenylosuccinate lyase (EC 4.3.2.2, adenylosuccinase, ADL) and 4 novel genes (one with probable rabgap domains and Src homology domain 3). Contains ESTs, STSs, GSSs and a putative Cpg island, complete sequence.
 AL022238
 AL022238.1 GI:4176442
 HTG: Adenylosuccinate lyase; ADL; Cpg Island.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE Direct Submission
 AUTHORS Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, UK
 TITLE CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk
 JOURNAL requests: clonerequest@sanger.ac.uk
 1 (bases 1 to 184180)

COMMENT This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the validation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP5-1042K10 is from the library RPCI-5 constructed at the Roswell Park Cancer

FEATURES

source

Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>
 VECTOR: PCVPAC2
 IMPORTANT: This sequence is not the entire insert of clone RP5-1042K10. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
 The true left end of clone RP5-1042K10 is at 1 in this sequence. The true left end of clone RP4-591M18 is at 184081 in this sequence. The true right end of clone RP3-377F16 is at 39440 in this sequence. The start of this sequence overlaps with sequence 283847.

Location/Qualifiers

1. 184180
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="22"
 /map="913.1-13.2"
 /clone="RP5-1042K10"
 /clone_11b="RPCI-5"
 1. 66
 /note="Alu repeat: matches 234. .299 of consensus"
 complement(5. .477)
 /note="match: GSS: Em:AQ152699"
 69. .380
 /note="Alu repeat: matches 1. .309 of consensus"
 765. .912
 /note="MIR repeat: matches 61. .215 of consensus"
 913. .1209
 /note="Alu repeat: matches 1. .297 of consensus"
 1225. .1402
 /note="Alu repeat: matches 140. .312 of consensus"
 1419. .1459
 /note="MIR repeat: matches 32. .71 of consensus"
 1469. .1766
 /note="Alu repeat: matches 1. .299 of consensus"
 1786. .1903
 /note="FLAM_C repeat: matches 1. .132 of consensus"
 1967. .2589
 /note="match: STS: Em:BI4103"
 match: GSS: Em:BI4103"
 join(3190. .3233, 3378. .3561, 10740. .10877, 13045. .13221, 14733. .14879, 15130. .15225, 17511. .17735, 18170. .18240, 23318. .23457)
 /gene="DJ1042K10.4"
 join(<3190. .3233, 3378. .3561, 10740. .10877, 13045. .13221, 14733. .14879, 15130. .15225, 17511. .17735, 18170. .18240, 23318. .23457)
 /gene="DJ1042K10.4"
 /note="supported by GENSCAN and GENES
 match: ESTs: Em:AA615306 Em:AA065323 Em:W80019 Em:N75651 Em:AA747718 Em:AA354286"
 /codon_start=3
 /evidence="not experimental"
 /product="DJ1042K10.4 (novel protein)"
 /protein_id="CA18263.1"
 /db_xref="GI:4176443"
 /db_xref="SPTREMBL:O95512"
 /translation="RKISOAVROODEOILAMVSALQOQOQOQOQOQPMKHSHPYV
 PKPILDNVPAALNGLDLOTKGIPGYSGSGSGADYGVGKAGTSSRFKQWT
 SMKEGLPSVAIOEAMHKNKAIVAGKTRGSPYNOFIIIGDITGLGHTGPAQGSWLP
 AKSPPTNKIGTSSNSNAWSPPEOPGPKGIIONIDPESDPVTVGSGVGTATSPYD
 TDHOLRLNRTTGSNSLNTSLSPRAWPKYSASDSTFNVHSTSAKFPDYKTSWSDPI
 GINPHILNKKMKKNHNTSSRNTPPLRPDPGLTNKRPSSPMSSTPAPRSVRCMGCTDSDRL
 ASASTWSGGSVPSYVWVLTLPQIDGSTLTITCMOHGLTLFHLALVTGTLIRX
 STKQRAKAQRTALH"
 3607. .3797
 /note="MIR repeat: matches 31. .226 of consensus"
 3826. .3970
 /note="L1MB8 repeat: matches 5746. .5894 of consensus"
 3885. .4574
 /note="match: GSS: Em:A0627708"
 3904. .4459
 misc_feature

```

repeat_region /note="match: GSS: Em: A0549653"
3974..3998 /note="MER46A repeat: matches 149..167 of consensus"
repeat_region 3999..4074 /note="Tiger4(Zomb1) repeat: matches 1..82 of consensus"
4581..4879 /note="Alusg repeat: matches 1..299 of consensus"
4986..5106 /note="Charlie2 repeat: matches 34..149 of consensus"
5266..5553 /note="Alusx repeat: matches 1..312 of consensus"
5554..5859 /note="Charlie2 repeat: matches 2067..2378 of consensus"
5860..6145 /note="Alusx repeat: matches 3..290 of consensus"
6146..7010 /note="Charlie2 repeat: matches 2378..3214 of consensus"
7441..7543 /note="L2 repeat: matches 2609..2710 of consensus"
8063..8176 /note="57 copies 2 mer tt 64 conserved"
8239..8547 /note="AlusA8 repeat: matches 1..308 of consensus"
8932..9225 /note="Alusg repeat: matches 1..294 of consensus"
/note="match: GSS: Em: A0170974"
9964..10284 /note="L2 repeat: matches 2062..2388 of consensus"
/note="match: GSS: Em: A0793345"
12506..13425 /note="match: GSS: Em: A073124"
12826..12925 /note="MIR repeat: matches 33..134 of consensus"
13772..14070 /note="Alusg repeat: matches 1..298 of consensus"
/note="match: GSS: Em: A0523109"
15998..16220 /note="MIR repeat: matches 51..262 of consensus"
18301..18603 /note="Alus repeat: matches 1..303 of consensus"
18627..18860 /note="Alus repeat: matches 16..287 of consensus"
19526..19819 /note="Alus repeat: matches 1..294 of consensus"
20494..20787 /note="match: GSS: Em: A0002626"
21016..21309 /note="Alusx repeat: matches 1..293 of consensus"
21447..21511 /note="Alu repeat: matches 244..298 of consensus"
21585..21640 /note="MER44A repeat: matches 192..243 of consensus"
21641..21950 /note="Alu repeat: matches 2..310 of consensus"
21951..22119 /note="MER44A repeat: matches 2..192 of consensus"
22428..22726 /note="Alusg repeat: matches 1..299 of consensus"
/note="match: GSS: Em: A056569"
23734..24027 /note="AlusB8 repeat: matches 1..303 of consensus"
24423..24690 /note="L1M1 repeat: matches 6053..6324 of consensus"
24876..24915 /note="MER2A repeat: matches 119..159 of consensus"
/note="match: GSS: Em: A069725"

```

```

repeat_region 25992...26035
repeat_region 26107..26176 /note="35 copies 2 mer gt 68 conserved"
misc_feature /note="match: STS: Em: H63387"
28689..29182 /note="match: GSS: Em: B14190"
29245..29777 /note="match: GSS: Em: A0715590"
29258..30130 /note="match: GSS: Em: A0738705"
misc_feature

Query Match 1.7% Score 22: DB 9: Length 184180;
Best Local Similarity 100.0% Pred. No. 5.2;
Matches 22: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Qy 1129 TTTGATATCTCACTAAGTAA 1150
|||||TTTTTTTTTTTTTTTTTTT
Db 136023 TTTGATATCTCACTAAGTAA 136044

RESULT 29
AC084825/2 216425 bp DNA HTG 03-FEB-2001
LOCUS Mus musculus clone RP23-114F14, WORKING DRAFT SEQUENCE, 27
DEFINITION
AC084825.2 GI:12658007
AC084825 HTG: HTGS_PHAHEL: HTGS_DRAFT.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 216425)
McCombie, W.R., Baker, J.P., Bahret, A., Bal, H., Balija, V.,
Dedhia, N.N., de la Bastide, M., Huang, E.N., King, L., Kirchoff, K.A.,
Miller, B., Nascimento, L.U., O'Shaughnessy, A.L., Preston, R.R.,
Rodriguez, S., Santos, L., Shah, R.S., Spiegel, L.A., Toth, K., Vill, M.D.
and Zlatavern, T.
TITLE Mouse Genomic Sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 216425)
AUTHORS McCombie, W.R.
TITLE Direct Submission
JOURNAL Submitted (22-NOV-2000) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
COMMENT On Feb 3, 2001 this sequence version replaced gi:11276150.
Genome Center
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Center code: GSHL
Web site: http://www.cshl.org/genseq
Contact: mccombie@cshl.org
Project Information
Center project name: RP23-114F14
Center clone name: RP23-114F14
* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 24308: contig of 24308 bp in length
* 24309 24595: gap of unknown length
* 24596 42729: contig of 18134 bp in length
* 42730 43016: gap of unknown length
* 43017 56375: contig of 13359 bp in length
* 56376 56662: gap of unknown length

```

```

* 56663 69379: contig of 12717 bp in length
* 69380 69666: gap of unknown length
* 69667 81976: contig of 12310 bp in length
* 81977 82262: gap of unknown length
* 82263 93885: contig of 11623 bp in length
* 93886 94171: gap of unknown length
* 94172 105626: contig of 11455 bp in length
* 105627 105912: gap of unknown length
* 105913 116480: contig of 10568 bp in length
* 116481 116766: gap of unknown length
* 116767 125888: contig of 9122 bp in length
* 125889 126174: gap of unknown length
* 126175 134777: contig of 8603 bp in length
* 134778 135063: gap of unknown length
* 135064 143634: contig of 8571 bp in length
* 143635 143920: gap of unknown length
* 143921 150477: contig of 6557 bp in length
* 150478 150763: gap of unknown length
* 150764 157228: contig of 6465 bp in length
* 157229 157515: gap of unknown length
* 157516 163965: contig of 6451 bp in length
* 163966 164251: gap of unknown length
* 164252 170636: contig of 6385 bp in length
* 170637 170932: gap of unknown length
* 170933 177237: contig of 6315 bp in length
* 177238 177523: gap of unknown length
* 177524 182030: contig of 4507 bp in length
* 182031 182316: gap of unknown length
* 182317 186721: contig of 4405 bp in length
* 186722 187007: gap of unknown length
* 187008 191332: contig of 4325 bp in length
* 191333 191618: gap of unknown length
* 191619 193400: contig of 3722 bp in length
* 193401 195626: gap of unknown length
* 195627 199195: contig of 3569 bp in length
* 199196 199481: gap of unknown length
* 199482 203031: contig of 3550 bp in length
* 203032 203317: gap of unknown length
* 203318 206200: contig of 2883 bp in length
* 206201 206486: gap of unknown length
* 206487 209217: contig of 2731 bp in length
* 209218 209503: gap of unknown length
* 209504 212144: contig of 2641 bp in length
* 212145 212430: gap of unknown length
* 212431 214894: contig of 2464 bp in length
* 214895 215180: gap of unknown length
* 215181 215181: contig of 1245 bp in length.

```

FEATURES
source
1. 216425
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-114F14"

BASE COUNT 57862 a 49581 c 47418 g 54051 t 7513 others
ORIGIN

Query Match 1.7%; Score 22; DB 2; Length 216425;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1250 TGTTCGGAATTAATTTATT 1271
|||||
Db 60664 TGTTCGGAATTAATTTATT 60643

RESULT 30
AC073945/c 232710 bp DNA HTG 10-AUG-2001
LOCUS AC073945
DEFINITION Mus musculus clone RP23-17le18, WORKING DRAFT SEQUENCE, 16
unordered pieces.
AC073945
AC073945.18 GI:14861741
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 232710)
AUTHORS Song, L., Jiang, X. and Roe, B.A.
TITLE Mus musculus BAC clone rp23-17le18
JOURNAL Unpublished
2 (bases 1 to 232710)
AUTHORS Song, L., Jiang, X. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (07-JUL-2000) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

COMMENT
OK Jul 18, 2001 this sequence version replaced gi:14647276.
NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

* 1. 2225: contig of 2225 bp in length
* 2226 2325: gap of unknown length
* 2326 4369: contig of 2044 bp in length
* 4370 4469: gap of unknown length
* 4470 6609: contig of 2140 bp in length
* 6610 6710: gap of unknown length
* 6710 8912: contig of 2203 bp in length
* 8913 9012: gap of unknown length
* 9013 11729: contig of 2717 bp in length
* 11730 11829: gap of unknown length
* 11830 15754: contig of 3925 bp in length
* 15755 18726: gap of unknown length
* 18727 18826: contig of 2872 bp in length
* 18827 21020: gap of unknown length
* 21021 21120: contig of 2194 bp in length
* 21121 21121: gap of unknown length
* 21122 23198: contig of 2078 bp in length
* 23199 23298: gap of unknown length
* 23299 27969: contig of 4671 bp in length
* 27970 28069: gap of unknown length
* 28070 32135: contig of 4066 bp in length
* 32136 32235: gap of unknown length
* 32236 35506: contig of 3271 bp in length
* 35507 35606: gap of unknown length
* 35607 35607: gap of unknown length
* 35608 59059: contig of 23453 bp in length
* 59060 59159: gap of unknown length
* 59160 102137: contig of 42978 bp in length
* 102138 102237: gap of unknown length
* 102238 163523: contig of 61286 bp in length
* 163524 163624: gap of unknown length
* 163624 232710: contig of 69087 bp in length.

```

FEATURES
source
1. 232710
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-17le18"

BASE COUNT 73185 a 44003 c 43046 g 70959 t 1517 others
ORIGIN

Query Match 1.7%; Score 22; DB 2; Length 232710;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1125 CATTTCATATCTCACTAAG 1146
|||||
Db 79460 CATTTCATATCTCACTAAG 79439

RESULT 31

```
HS229041      HS229041      323000 bp      DNA      PRI      22-MAY-1998
LOCUS      Homo sapiens 959 kb contig between AML1 and CBRL on chromosome
DEFINITION      21q22: segment 1/3.
ACCESSION      AJ229041 AJ229040
VERSION      AJ229041.1 GI:3153101
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 323000)
AUTHORS      Bleischmidt,K., Rump,A., Nordsiek,G., Drescher,B., Weber,J. and
               Rosenthal,A.
TITLE      Sequencing and Analysis of 960 kb between AML1 and CBRL on
               Chromosome 21q22.2
JOURNAL      Unpublished
               2 (bases 1 to 323000)
REFERENCE      Bleischmidt,K., Rump,A., Nordsiek,G., Drescher,B., Weber,J.,
               Schattevoy,R., Rosenthal,A., Yaspo,M.L., Osoegawa,K. and Soeda,E.
AUTHORS      Direct Submission
TITLE      Submitted (07-MAY-1998) Karin Bleischmidt, Institute of Molecular
               Biotechnology, Dept. of Genome Analysis, Beutenbergstrasse 11,
               07745 Jena, GERMANY
JOURNAL      Segment entries: AJ229041, AJ229042, AJ229043
               Complete contig available via ftp:
               ftp.ebi.ac.uk/pub/databases/emb1/genomes/hsap1ens/contigs/qj229040/

COMMENT
FEATURES
    source      1. Location/Qualifiers
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="21"
               /map="q22"
               2. .46430
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="Cosmid Q98A3 (AF020802)"
               160. .289
               /rpt_family="FLAM_A"
               /evidence=not_experimental
               complement(917. .967)
               /rpt_family="MLT1B"
               /evidence=not_experimental
               complement(1046. .1348)
               /rpt_family="AluIo"
               /evidence=not_experimental
               complement(1349. .1559)
               /rpt_family="MLT1B"
               /evidence=not_experimental
               complement(2055. .2418)
               /rpt_family="MLT1A1"
               /evidence=not_experimental
               complement(2422. .2560)
               /rpt_family="MIR"
               /evidence=not_experimental
               complement(2579. .2870)
               /rpt_family="AluSx"
               /evidence=not_experimental
               complement(2878. .2964)
               /rpt_family="MIR"
               /evidence=not_experimental
               complement(2981. .3198)
               /rpt_family="MIR"
               /evidence=not_experimental
               complement(3702. .3821)
               /note="GRAIL score = 60.000%, comment = good"
               /evidence=not_experimental
               complement(3770. .3826)
               /note="Genscan, score = 9.66%, comment = Initial_exon 57
               bp frame: 1 phase: 0"
               /evidence=not_experimental
               complement(4584. .4601)
               /evidence=not_experimental
               complement(4671. .4730)
               /note="GRAIL score = 84.000%, comment = excellent"
               /evidence=not_experimental
               complement(4999. .5277)
               /note="GRAIL score = 43.000%, comment = marginal"
               /evidence=not_experimental
               complement(5023. .5134)
               /note="Xpound exon prediction, score = 85% (0%)"
               /evidence=not_experimental
               complement(5730. .5823)
               /note="GRAIL score = 84.000%, comment = excellent shadow"
               /evidence=not_experimental
               complement(5995. .6057)
               /rpt_family="MER4_Internal"
               /evidence=not_experimental
               complement(6560. .6647)
               /rpt_family="L2"
               /evidence=not_experimental
               complement(6719. .6898)
               /note="Genscan, score = 1.84%, comment = Internal_exon 180
               bp frame: 1 phase: 0"
               /evidence=not_experimental
               complement(8517. .8768)
               /rpt_family="MLT2G"
               /evidence=not_experimental
               complement(9065. .9199)
               /rpt_family="MER5B"
               /evidence=not_experimental
               complement(9679. .10211)
               /rpt_family="LIMD3"
               /evidence=not_experimental
               complement(10721. .11087)
               /rpt_family="MLT1B"
               /evidence=not_experimental
               complement(11094. .11392)
               /rpt_family="AluSg"
               /evidence=not_experimental
               complement(11856. .12044)
               /rpt_family="L2"
               /evidence=not_experimental
               complement(12225. .12338)
               /rpt_family="L2"
               /evidence=not_experimental
               complement(12294. .12390)
               /note="GRAIL score = 60.000%, comment = good"
               /evidence=not_experimental
               complement(14078. .14380)
               /rpt_family="AluSx"
               /evidence=not_experimental
               complement(14490. .14717)
               /note="Genscan, score = 3.65%, comment = Terminal_exon 228
               bp frame: 2 phase: 0"
               /evidence=not_experimental
               complement(15634. .15678)
               /rpt_family="L2"
               /evidence=not_experimental
               complement(15687. .15846)
               /note="GRAIL score = 65.000%, comment = good"
               /evidence=not_experimental
               complement(16160. .16454)
               /rpt_family="AluSp"
               /evidence=not_experimental
               complement(17435. .17595)
               /rpt_family="AluIb"
               /evidence=not_experimental
               complement(17657. .17936)
               /rpt_family="AluSg"
               /evidence=not_experimental
               complement(18646. .18944)
               /rpt_family="AluSc"
               /evidence=not_experimental
```

```

repeat_region complement(20104..20412)
                /rpt_family="AluIo"
                /evidence=not_experimental
source          20869..61948
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="Cosmid c103C0352 (AF020802)"
misc_difference 21121..21145
                /note="Deletion in Q98A3"
                /replace=""
exon            21541..21656
                /note="GRAIL, score = 86.000%, comment = excellent
                Genscan, score = 7.63%, comment = Initial-exon 116 bp
                frame: 0 phase: 2"
                /evidence=not_experimental
misc_difference 21782
                /note="g to A in Q98A3"
                /replace="a"
exon            complement(22029..22079)
                /note="GRAIL, score = 53.000%, comment = good shadow"
                /evidence=not_experimental
exon            complement(22957..23069)
                /note="GRAIL, score = 49.000%, comment = marginal shadow"
                /evidence=not_experimental
repeat_region  complement(23334..23382)
                /rpt_family="L2"
                /evidence=not_experimental
repeat_region  23386..23680
                /rpt_family="AluSx"
                /evidence=not_experimental
repeat_region  complement(23681..23784)
                /rpt_family="L2"
                /evidence=not_experimental
repeat_region  complement(23944..24252)
                /rpt_family="LMC1"
                /evidence=not_experimental
repeat_region  complement(25717..25835)
                /rpt_family="MER5B"
                /evidence=not_experimental
repeat_region  complement(25893..26448)
                /rpt_family="L1PA12"
                /evidence=not_experimental
exon            complement(26017..26086)
                /note="GRAIL, score = 42.000%, comment = marginal shadow"
                /evidence=not_experimental
exon            26475..26549
                /note="GRAIL, score = 75.000%, comment = excellent"
                /evidence=not_experimental
misc_difference 26505
                /note="G to A in Q98A3"
                /replace="a"
exon            27130..27258
                /note="Genscan, score = 2.49%, comment = Internal-exon 129
                bp frame: 1 phase: 0"
                /evidence=not_experimental
Query Match    1.7%; Score 22; DB 9; Length 323000;
                Best Local Similarity 100.0%; Pred. No. 5.1;
                Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 681 GAAACCGAAGAAATGAGTTTGC 702
DB 57236 GAAACCGAAGAAATGAGTTTGC 57257

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
1 (sites) Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., and Schudy,A.
The DNA sequence of human chromosome 21. The chromosome 21 mapping
and sequencing consortium
Nature 405 (6784), 311-319 (2000)
2 (bases 1 to 340000)
JOURNAL MEDLINE
20289799
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Mitsuyama,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G.,
Horisch,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Kieselmann,L., Dagnid,E., Wehrmeyer,S., Borzym,K.,
Gardiner,K., Nizetic,D., Francis,F., Lehnach,H., Reinhardt,R. and
Yaspo,M.L.
Direct Submission
Submitted (10-APR-2000) to the DDBJ/EMBL/Genbank databases. The
Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic
Sciences Center, Human Genome Research Group * Institute of
Molecular Biotechnology, Genome Biology * Keio University School
of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome
Analysis * Max-Planck Institute for Molecular Genetics (addresses
see below)
On May 30, 2000 this sequence version replaced gi:7717338.
The chromosome 21 mapping and sequencing consortium consisting of
* RIKEN Genomic Sciences Center, Human Genome Research Group, *
Sagamihara 228-8555, Japan,
* e-mail: hattori@gsf.riken.go.jp/
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, *
Beutenbergstrasse 11, D-07745 Jena, Germany,
* e-mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Molecular Biology, * Tokyo
160-8582, Japan,
* e-mail: nshimizudmb-med.keio.ac.jp/
* URL: http://www.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e-mail:
info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
Innestrasse 73, D-14195 Berlin, Germany,
* e-mail: info-chr21@molgen.mpg.de/
* URL: http://chr21.rz-berlin.mpg.de/
AL163269: Submitted (10-APR-2000).
Location/Qualifiers
1..340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/Chromosome="21"
<1..53351
/note="Accession No. AF015720"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/Chromosome="21"
/map="21q22.1"

```

```

source
AP001724 340000 bp DNA PRI 30-MAY-2000
DEFINITION Homo sapiens genomic DNA, chromosome 21q, section 68/105.
ACCESSION AP001724 AL163269 BA000005
VERSIONS AP001724.1 GI:7768721
KEYWORDS
SOURCE Homo sapiens DNA.

```

```

/clone="p169k17, 5' partial"
/clone_id="RPC11,3-5 PAC library"
851. .1099
/repeat_region
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(2197. .2499)
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(2741. .3129)
/note="THE1C"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
3756. .3945
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
4224. .4272
/note="(GA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(4692. .5067)
/note="THE1C"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
5224. .5728
/note="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
5742. .6191
/note="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
6275. .6577
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(9217. .9753)
/note="MLT2D"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
complement(10700. .10816)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(11045. .11141)
/note="MER5B"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
12481. .13054
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
13090. .13456
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(13457. .13859)
/note="LIPa8"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
13860. .13927
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
14377. .14565
/note="MER5A"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
complement(14684. .14770)
/note="MLT1J"
/rpt_family="LTR/MaLR"

```

```

/repeat_region
/rpt_type=DISPERSED
complement(14845. .15169)
/note="MLT1J"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
17813. .17917
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
18095. .18315
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
18611. .18708
/note="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
18760. .18960
/note="(TAA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
18775. .18893
/note="FLAM_C"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
18833. .63876
/note="Accession No. AF020803"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="Q37D2"
/clone_lib="L121NCO2-Q Cosmid library"
18894. .18960
/note="(TAA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(19309. .19427)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
20358. .20646
/note="LIME3A"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
20647. .20954
/note="AluJo"
/rpt_family="SINE/Alu"

Query Match 1.7%: Score 22; DB 9; Length 340000;
Best local similarity 100.0%: Pred. No. 5.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 681 GAAACGAGAAATGAGTTTC 702
|||||
Db 125791 GAAACGAGAAATGAGTTTC 125770

RESULT 33
CNS061A0 958 bp DNA STS 10-JAN-2001
LOCUS T7 end of clone AS0A019A04 of library AS0A from strain CLIB 533
DEFINITION of Saccharomyces bayanus, sequence tagged site.
ACCESSION A1400030
VERSION A1400030.1 GI:12155369
KEYWORDS STS.
SOURCE Saccharomyces bayanus.
ORGANISM Saccharomyces bayanus.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 958)
AUTHORS Bon,E., Neuvéglise,C., Casaregola,S., Attienave,F., Wincker,P.,
Aigle,M. and Durrens,P.

```

TITLE Genomic Exploration of the Hemiascomycetous Yeasts: 5.
JOURNAL Saccharomyces bayanus var. uvarum
PUBMED FEBS Lett. 487 (1), 37-41 (2000)
REFERENCE 11152880
AUTHORS 2 (bases 1 to 958)
Souchet, J. L., Aigle, M., Artiguenave, F., Blandin, G.,
Bollotin-Pikuhara, M., Bon, E., Brottier, P., Casaregola, S.,
de-Montigny, D., Dujon, B., Durans, P., Lepingle, A., Llorente, B.,
Maupertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Pottier, S.,
Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
Winkler, P. and Weissenbach, J.
TITLE Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
yeast species for molecular evolution studies(1)
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
PUBMED 11152876
REFERENCE 3 (bases 1 to 958)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This STS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
source
Location/Qualifiers
1..958
/organism="Saccharomyces bayanus"
/strain="Clb 533"
/variety="uvarum"
/db_xref="taxon:4931"
/clone="AS0A019A04"
/clone_1lb="AS0A0A"
/note="end : T7"
misc-feature
c23..>462
/note="similar to Saccharomyces cerevisiae ORF YBR011c [
IPPI : Inorganic pyrophosphatase, cytoplasmic]
1 putative frameshift(s)"
/evidence=not experimental
BASE COUNT 317 a 183 c 168 g 287 t 3 others
ORIGIN

Query Match 1.6%; Score 21; DB 11; Length 958;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 545 GACTGGAAGTCATTCGCATT 565
|||||
Db 58 GACTGGAAGTCATTCGCATT 78

RESULT 34
LOCUS SCF91 39525 bp DNA BCT 23-NOV-1999
DEFINITION Streptomyces coelicolor cosmid F91.
ACCESSION AL132973
VERSION 1
KEYWORDS
2-hydroxyacid-family dehydrogenase; anti-sigma factor antagonist;
binding protein dependent transport protein; catalase; catB;
Deor-family transcriptional regulator; gas vesicle synthesis
protein; glycerol-3-phosphate dehydrogenase; gvpa; gvpr; gvpg;
gvpj; gvpk; gvpl; gvpo; gvps; membrane protein; methyltransferase;
secreted cellulose-binding protein; secreted chitosanase; secreted
endo-1,4-beta-xylanase; short chain oxidoreductase; soluble-binding
lipoprotein; tellurite resistance protein; terd; tetr-family
transcriptional regulator; transmembrane efflux protein; xysa;

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

zinc-containing dehydrogenase.
Streptomyces coelicolor A3(2).
Streptomyces coelicolor A3(2)
Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:
Actinomycetales; Streptomycetaceae; Streptomycetes.
1 (bases 1 to 39525)
Redenbach, M., Kleser, H.M., Denapalte, D., Eichner, A., Cullum, J.,
Knaash, H. and Hopwood, D.A.
A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
97000351
2 (bases 1 to 39525)
Oliver, K. and Harris, D.
Unpublished
3 (bases 1 to 39525)
Thomson, N.R., Parkhill, J., Barrell, B.G. and Rajandream, M.A.
Direct Submission
Submitted (12-NOV-1999) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>)
CDS are numbered using the following system eg SC787.01c. SC (S.
coelicolor), 787 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in codons
using a specially developed Hidden Markov Model (Krogh et al.,
Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
program of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nih.go.jp/>
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtg, ttg or att) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring subclones.
Cosmid F91 overlaps with cosmid STF15 on the AseI-F genomic
restriction fragment.

FEATURES
source
Location/Qualifiers
1..39525
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid F91"
1..247
/gene="terd"
complement(1..228)
/note="nominal overlap with cosmid STF56"
1..247
/gene="terd"
/note="SCF91.01, terd, tellurite resistance protein,
partial CDS, len: > 81 aa. Highly similar to many
tellurite resistance proteins including: Serratia
marcescens SW:TERD_SRRM1(EMBL:138824) tellurium resistance
protein Terd (192 aa), fasta scores opt: 398 z-score:

	misc.feature						
RBS	gene	/note="P800216 Sugar transport proteins signature 1"					
CDS		/gene="+SCF91_04c."					
		/complement(5163_.5167)					
		/complement(5266_.6279)					
		/gene="+SCF91_05c."					
		/complement(5266_.6279)					
		/gene="+SCF91_05c."					
		/note="+SCF91_05c, possible zinc-containing dehydrogenase,					
		len : 337 aa. Similar to many dehydrogenases including:					
		Alicalligenes eutrophus SW:ADH_ALCEU[EMBL:j03362] alcohol					
		dehydrogenase [EC 1.1.1.1] (366 aa), fasta scores opt :					
		272 z-score: 296.1 E() : 4e-09 27.0% identity in 371 aa					
		overlap and Streptomyces coelicolor TR:O53927[EMBL:X62373]					
		hydoxyacyl-CoA dehydrogenase (339 aa), fasta scores opt :					
		308 z-score: 335.0 E() : 2.8e-11 31.8% identity in 333 aa					
		overlap. Contains a Prosite hit to P500059 zinc-containing					
		alcohol dehydrogenases signature and a Pfam match to entry					
		Pf00107 adh_zinc, zinc-binding dehydrogenases."					
		//codon_start=1					
		/transl_table=1					
		/product="putative zinc-containing dehydrogenase."					
		/protein_id="CAB61162.1"					
		/db_xref="GI:6434715"					
		/translation="MROMLAARLHVPSRTLRLEVPDQPGGVLVKKVAAGCLSD					
		VHLIDGLTLPFLRLRGDTVTTLGHEYSYTAGTGACGYTWKSTEQRYALTEBRDCVTTT					
		RCVDYDGMAEYAICADAMTPLEDIAIPFGDAITPDASTPWCAIETGEVERPAEEVV					
		GWGVGIGLYAHAVOLLRAIGCAPVAAVDPSPVRERAILAGADAIDSAPFEROKGVN					
		AATGGACLTAADFAPGRPVREQNVSLAPKRGLVLGLNDKPPLTVTDGTRESVLOOR					
Query Match	Best Local Similarity	1.6%	Score 21;	DB 1;	Length 39525;		
Matches	Conservative	0;	Mismatches	0;	Indels	0; Gaps 0;	
Oy	95 GGCTTCAGCACCGAGGACGC 115 Db 33904 GGCTTCAGCACCGAGGACGC 33924						
RESULT 35 AF020801/c LOCUS DEFINITION Homo sapiens chromosome 21 clone cosmid c102A0977 map 21q22.2.	AF020801 complete sequence.	DNA Pri	05-MAY-2001				
ACCESSION VERSION APO20801 KEYWORDS SOURCE HOMO ORGANISM Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 44548) Hattori,M., Fujiyama,A., Taylor,T.D., Matanabe,H., Yada,T., Park,H.-S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.-K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Bleeschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Paterson,W., Reichwald,K., Rump,A.J., Schillhaber,M.B., Schudy,A., Zimmermann,N., Rosenthal,A., Kudoh,J., Kawasaki,K., Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyma,S., Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G., Antonaracher,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,						

annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Ch13> Rpl1-218A18 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: PBAC3.6

IMPORTANT: This sequence is not the entire insert of clone Rpl1-218A18. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone Rpl1-270H22 is at 60060 in this sequence. The true right end of clone Rpl1-380B4 is at 100 in this sequence.

FEATURES

```
source
location/Qualifiers
1..60159
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone_lib="RPCI-11.1"
/clone_id="RPCI-11.1"
2288..2579
/note="Aluub repeat: matches 1..288 of consensus"
repeat_region
3709..4038
/note="MLTJ repeat: matches 164..508 of consensus"
4377..4566
/note="LTR16C repeat: matches 199..387 of consensus"
4570..4689
/note="MLTJ repeat: matches 36..158 of consensus"
5281..5594
/note="Aluuo repeat: matches 5..309 of consensus"
6203..6268
/note="MLTJ repeat: matches 446..513 of consensus"
6333..6416
/note="MLTJ repeat: matches 273..354 of consensus"
6676..6832
/note="MIR repeat: matches 88..262 of consensus"
7355..7551
/note="MIR repeat: matches 47..254 of consensus"
7640..8678
/note="LIME1 repeat: matches 5177..6158 of consensus"
8679..8972
/note="Aluuo repeat: matches 1..300 of consensus"
8973..9682
/note="LIME1 repeat: matches 4507..5177 of consensus"
9683..9972
/note="Aluub repeat: matches 1..296 of consensus"
9973..10663
/note="LIME1 repeat: matches 3855..4507 of consensus"
10664..11161
/note="MLTJ repeat: matches 1..503 of consensus"
11162..12453
/note="LIME1 repeat: matches 2245..3855 of consensus"
12442..14102
/note="LIRP7 repeat: matches 4477..6140 of consensus"
15771..15816
/note="L2 repeat: matches 2599..2644 of consensus"
16213..16388
/note="MERS3 repeat: matches 18..189 of consensus"
17153..17217
/note="L2 repeat: matches 2644..2708 of consensus"
17232..17815
/note="LIRP1 repeat: matches 5522..6154 of consensus"
18064..18115
/note="FAM repeat: matches 122..173 of consensus"
20102..20282
/note="L2 repeat: matches 2370..2557 of consensus"
20408..21011
/note="LIM4 repeat: matches 225..841 of consensus"
21222..21316
/note="MERSA repeat: matches 7..103 of consensus"
21369..21670
/note="AluSg repeat: matches 1..300 of consensus"
21730..21954
/note="MLTJ repeat: matches 270..494 of consensus"
21978..22097
/note="L2 repeat: matches 2589..2710 of consensus"
23357..23786
/note="L2 repeat: matches 2260..2694 of consensus"
23831..24397
/note="LIMB2 repeat: matches 4838..5421 of consensus"
24398..24692
/note="AluSg repeat: matches 1..295 of consensus"
24693..25474
/note="LIMB2 repeat: matches 5421..6168 of consensus"
26438..26696
/note="AluSg repeat: matches 36..303 of consensus"
26804..26857
/note="MERSB repeat: matches 1..54 of consensus"
26881..28220
/note="L2 repeat: matches 1267..2708 of consensus"
27519..27539
/note="Sequence from clone PCR only"
28272..28513
/note="LIMB2 repeat: matches 5803..6046 of consensus"
31634..31677
/note="L2 repeat: matches 70..238 of consensus"
32513..32678
/note="MIR repeat: matches 70..238 of consensus"
32947..33046
/note="L2 repeat: matches 2606..2707 of consensus"
33348..33636
/note="MERSA repeat: matches 97..183 of consensus"
34078..34506
/note="L2 repeat: matches 211..2750 of consensus"
34591..34654
/note="L2 repeat: matches 211..2750 of consensus"
34656..34977
/note="L2 repeat: matches 211..2750 of consensus"
35075..35251
/note="MIR repeat: matches 27..390 of consensus"
35582..36885
/note="LIRP1 repeat: matches 5974..6154 of consensus"
37331..37625
/note="AluSg repeat: matches 1..302 of consensus"
37776..37964
/note="AluSg repeat: matches 1..300 of consensus"
38141..38237
/note="MERSA repeat: matches 7..189 of consensus"
39030..39175
/note="MIR repeat: matches 36..147 of consensus"
39612..40032
/note="HERVL repeat: matches 4853..4999 of consensus"
40655..41314
/note="HERVL repeat: matches 1..448 of consensus"
41321..41470
/note="HERVL repeat: matches 4102..4787 of consensus"
41862..41905
/note="Charliel repeat: matches 2610..2761 of consensus"
42378..42680
/note="L2 repeat: matches 2599..2644 of consensus"
44890..45108
/note="AluSx repeat: matches 1..300 of consensus"
45109..45280
/note="LIME1 repeat: matches 6099..6320 of consensus"
45292..46245
/note="FAM repeat: matches 2..175 of consensus"
45292..46245
/note="LIRP3 repeat: matches 5191..6144 of consensus"
46251..46325
/note="FAM repeat: matches 52..126 of consensus"
```


exon /number=5
 complement(18688..18784)
 /gene="AT4935790"
 intron /number=6
 complement(18785..19340)
 /number=6
 complement(19341..19633)
 /gene="AT4935790"
 exon /number=7
 complement(19634..19909)
 /number=7
 complement(19910..20110)
 /gene="AT4935790"
 intron /number=8
 complement(20111..20249)
 /number=8
 complement(20250..20399)
 /gene="AT4935790"
 exon /number=9
 complement(20400..20514)
 /number=9
 complement(20515..20663)
 /gene="AT4935790"
 intron /number=10
 complement(20664..20871)
 /number=10
 complement(20872..21019)
 /gene="AT4935790"
 exon /number=11
 complement(21020..21160)
 /number=11
 complement(21161..21229)
 /gene="AT4935790"
 intron /number=12
 complement(21230..21318)
 /number=12
 complement(21319..21624)
 /gene="AT4935790"
 exon /number=13
 complement(21625..21714)
 /number=13
 complement(21715..21909)
 /gene="AT4935790"
 intron /number=14
 complement(21910..22186)
 /number=14
 complement(22187..22336)
 /gene="AT4935790"
 exon /number=15
 complement(22337..22496)
 /number=15
 complement(22497..23438)
 /gene="AT4935790"
 intron /number=16
 complement(23439..24846,24953..25024,25138..28199,
 28285..28543,28720..29006,29101..29436,29522..29663,
 29738..29958,30041..30202,30288..30446,30524..30808,
 30897..31181,31372..31455))
 /gene="AT4935800"
 gene /number=1
 complement(24678..24846,24953..25024,25138..28199,
 28285..28543,28720..29006,29101..29436,29522..29663,
 29738..29958,30041..30202,30288..30446,30524..30808,
 30897..31181,31372..31455))
 /note="In contrast to gb:S11960 gene contains an intron at
 position 29521-29437 which is present in similar proteins
 as well; the sequence differs at position 29520-29525 from
 gb:S11960; sequence stretch has been verified on BAC

Query Match 1.68; Score 21; DB 8; Length 99461;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 788 AAACGAATGCAAAAGCAATC 808
 Db 72542 AAACGAATGCAAAAGCAATC 72562

RESULT 39
 ATT19K4
 LOCUS
 DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone T19K4 (ESSAII
 Project).
 ACCESSION AL022373
 VERSION AL022373.1 GI:3036791
 KEYWORDS
 SOURCE
 ORGANISM
 thale cress.
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE
 1 (bases 1 to 106007)
 Bevan, M., Medler, H., Wambutt, R., Hohnselt, J., Jesse, T., Heljnen, L.,
 Vos, P., Mewes, H.W., Mayer, K. and Schueller, C.
 JOURNAL
 2 (bases 1 to 106007)
 unpublished
 REFERENCE
 2
 EU Arabidopsis sequencing project.
 AUTHORS
 Submitted (07-APR-1998) MIPS, at the Max-Planck-Institut fuer
 JOURNAL
 Biochemie, Am Klopfersplitz 18a, D-82152 Martinsried, FRG, E-mail:
 'schueller@mips.biochem.mpg.de or mayer@mips.biochem.mpg.de Project
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
 Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
 E-mail: michael.bevan@bbsrc.ac.uk
 FEATURES
 Location/Qualifiers
 1..106007
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="4"
 285..2210
 /gene="T19K4.10"
 285..374
 /gene="T19K4.10"
 /number=1
 join(285..374,483..718,803..1027,1117..1252,1351..1412,
 1870..1904,1987..2210)
 /gene="T19K4.10"
 /note="strong similarity to predicted protein A_TW017A05.8
 of Arabidopsis thaliana BAC TW017A05, Patchx:G2435517

Contains eukaryotic and viral aspartyl proteases active sites [VALDTGSDLFMV] and [ALFDVGTSTYL]."

/codon_start=1
/product="putative protein (fragment)"
/protein_id="CA18482.1"
/db_xref="GI:3036792"
/db_xref="SPTRMBL:O65625"
/translation="MVALDTGSDLFMVCDGCKAPTEGATVASEFELSIYKPVSTT
NKVYTNNSLCAQNNOCIGFTSCPYMYVSAQTSGLIMEDVMHLLTDDKPERV
EAYVFGGVOYSGSFLDIAPNGLFGLGEMKISVPSVLAREGLVADSFSCFGHDV
GRISFGDKGSDQETEPNLSHPNNTVYTRVRCFTLLIDDEFTALFDGTSFTYL
VDPMYTVSESAQDKRHSPOSRIPPEICYDMREKLVLAAMKFDCTDIEENVTAGTN
KTAAVAPMAAGIKTHNNSSELHKTNTITSKSSPNQISKTVDWSEFRFVFTLLPL
V"

375..482
/gene="T19K4.10"
/number=1
483..718
/gene="T19K4.10"
/number=2
719..802
/gene="T19K4.10"
/number=2
803..1027
/gene="T19K4.10"
/number=3
1028..1116
/gene="T19K4.10"
/number=3
1117..1252
/gene="T19K4.10"
/number=4
1253..1350
/gene="T19K4.10"
/number=4
1351..1412
/gene="T19K4.10"
/number=5
1413..1869
/gene="T19K4.10"
/number=5
1870..1904
/gene="T19K4.10"
/number=6
1905..1986
/gene="T19K4.10"
/number=6
1987..2210
/gene="T19K4.10"
/number=7
3922..6899
/gene="T19K4.20"
3922..4443
/gene="T19K4.20"
/number=1
join(3922..4443,4917..5406,5739..5809,6321..6392,
6636..6899)
/gene="T19K4.20"
/note="similarity to protein coded for by gene T12P5.5 in
Caenorhabditis elegans cosmid T12P5, Patchx:G2773225
contains EST gb:z26699"
/codon_start=1
/product="putative protein"
/protein_id="CA18483.1"
/db_xref="GI:3036793"
/db_xref="SPTRMBL:O65626"
/translation="MASATSNPSSSSMSPRISGNHGSPTASVAQSPRRRSROYSSP
WTQIVRGSEPIIAAAVAGPSSPQSRAPIEPIASVVAAPTAALVTEAAGDEKSE
ASSGODNAGKKPVKKRPSNGASEVPGVAGASWPALETTKAPSKSSDSLSKIGDV
PSSSSASSSVTQGIANAVAPAKQAGRANPTNHSRQSFQORAGSANGTV
SOPSAQSGFTEPSHNPSRPGCONKNGFASQNHGTEHNSRODSYRNQNGHSHG
RRNEHGNOMTEFORSPNGREGNAOSQGTAPAFVHRHSPPTVOPIDQPMAAQPFSSHIP
PFTDILNGPMNNEGFLRVYAGFKKVAELTDNIQQLVEALQNSPHVEVGDFTRKR

DMQNVLRNPRTGSGPOSVDRADAVAKRLGNLSVDSSADPIGSSSSQLOPTEALSD
DOOQSSSTAPVSNHNPADGANR"

4444..4916
/gene="T19K4.20"
/number=1
4917..5406
/gene="T19K4.20"
/number=2
5407..5738
/gene="T19K4.20"
/number=2
5739..5809
/gene="T19K4.20"
/number=3
5810..6320
/gene="T19K4.20"
/number=3
6321..6392
/gene="T19K4.20"
/number=4
6393..6635
/gene="T19K4.20"
/number=4
6636..6899
/gene="T19K4.20"
/number=5
11235..11969
/gene="T19K4.30"
11235..11969
/gene="T19K4.30"
/note="weak similarity to GBF4, Arabidopsis thaliana,
PAX:G403418
Contains bZIP transcription factors basic domain signature
[KRMIKRESARSRAR]"
/codon_start=1
/product="putative bZIP transcription factor"
/protein_id="CA18484.1"
/db_xref="GI:3036794"
/db_xref="SPTRMBL:O65627"
/translation="MLSSAKHQNRHRLSATNKNOTLRKVSISSSSPSSSSSSSTSS
SSPLPSODSQAKRSILVMEVMDINLASIHNRHSPHQRHNERFRQNHQNG
PNSIFQDFELKSLNQEPAPRTQOTGSPNGSDSTVLYSSPPPTVLSLNGAGF
EFLNODPIVTSNLSNLTHTHHLSNAHAFNFSFELVPSVSFGKRRGDSNGSGNRH
KRMIKRESARSRARKEQCSAPLPS"
11235..11969
/gene="T19K4.30"
/number=1
14550..17503
/gene="T19K4.40"
complement(14550..14672,14834..14891,15525..15919,
16020..16158,16200..16321,16597..16683,16777..17140,
17208..17503)
/gene="T19K4.40"
complement(join(14550..14672,14834..14891,15525..15919,
16020..16158,16200..16321,16597..16683,16777..17140,
17208..17503))
/gene="T19K4.40"
/note="similarity to predicted protein P29F11.3,
Caenorhabditis elegans, Patchx:E248390
contains EST gb:T21369"
/codon_start=1
/product="putative protein"
/protein_id="CA18485.1"
/db_xref="GI:3036795"
/db_xref="SPTRMBL:O65628"
/translation="MACNNSCGESGCGCDREKNGSKITVDVSGGNGHESVCYCKCN
APMTFGCGEDGRCADCEFNNTFGFLAVTSHAMITPSDNVLAFFSGSSSSSKVD
VRSLQVHEQIETALKNVEASRRSLPFGVGAFAVDETAAPALSTEMIDALEVR
VTSCLSPKMDLHVVEYESTFGSDSDARDRLKLDSDVDDGKRDLLHKKMLSL
QKVAENGYNRLVLGSGCTSLASVILRTATYKGRYSLSADQIHDAWMKYPITVPLAD

CVRLEITRCLDGEPEYSCITLISLVYSGGRKLTVELACSSGCGINDVSSFVAL
 LQENPSRECTIVRTAKLTPEYFNKIPEDSDNMAVDRKRLKKGDMTEA
 FCPICNPINRSDSELDPEEGESDVLAYACSSCRFDILPODSSSEFSLPD
 HMTSOVKHOKVDSOAYLRHYNTRKLMFEDFISKDGMMGRAPFLKSKSTIAQRSLIP
 ETGKKREKVRRELKIKALEW"

Intron
 complement(14673..14833)
 /number=1
 complement(14834..14891)
 /gene="T19K4.40"

exon
 /number=2
 complement(14892..15524)
 /number=2
 complement(15525..15919)
 /gene="T19K4.40"

Intron
 complement(15920..15919)
 /gene="T19K4.40"

Query Match
 Best Local Similarity 100.0%; Score 21; DB 8; Length 106007;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 15469 AAAACGATGCGAAAGGATC 808
 |||||||||||||||||||
 |||||||||||||||||||

QY 788 AAAACGATGCGAAAGGATC 808
 |||||||||||||||||||
 |||||||||||||||||||

Db 15469 AAAACGATGCGAAAGGATC 15489

RESULT 40
 AC008034/c
 LOCUS
 DEFINITION
 AC008034 120134 bp DNA HTG 17-JUL-2001
 Homo sapiens chromosome 3 clone CTB-134A23, WORKING DRAFT SEQUENCE,
 3 unordered places.
 AC008034
 AC008034.28 GI:14787124
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 120134)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
 Alstbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbata,J.,
 Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowler,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhaq,C.,
 Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carion,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Hollaway,C.,
 Hollins,B., Homsl,F., Howard,S., Huber,J., Hulik,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Koryah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,
 Lounseged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
 Meil,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nockenkwo,S.,
 Oguh,M., Okunonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
 Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,
 Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalton,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Unpublished
 2 (bases 1 to 120134)
 Morley,K.C.
 Direct Submission
 Submitted (13-JUL-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 17, 2001 this sequence version replaced gi:14547711.

 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu

 Project Information
 Center project name: HMHO
 Center clone name: CTB-134A23

 Summary Statistics
 Sequencing vector: M13, L08821
 Chemistry: Dye-terminator Big Dye: 30% of reads
 Chemistry: Dye-terminator Big Dye: 70% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 121889 bases at least 400
 Consensus quality: 124362 bases at least 030
 Consensus quality: 125920 bases at least 020
 Estimated insert size: 121022; sum-of-confids estimation
 Quality coverage: 10.7x in Q20 bases; sum-of-confids estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 77264: contig of 77264 bp in length
 * 77265 77364: gap of unknown length
 * 77365 117203: contig of 39839 bp in length
 * 117204 117303: gap of unknown length
 * 117304 120134: contig of 2831 bp in length.

FEATURES
 source
 1. 120134
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /clone="CTB-134A23"

BASE COUNT
 31243 a 29435 c 29477 g 29748 t 231 others

Query Match
 Best Local Similarity 100.0%; Score 21; DB 2; Length 120134;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 CTGTGCGAGCGGCGGCGAG 76
 |||||||||||||||||||
 |||||||||||||||||||

Db 99470 CTGTGCGAGCGGCGGCGAG 99450

RESULT 41
 AC011414/c
 LOCUS
 DEFINITION
 AC011414 Homo sapiens chromosome 5 clone CTB-83D3, complete sequence.
 AC011414.4 GI:14971175
 HTG.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 123224)
 TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
 AUTHORS Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 123224)
 TITLE DOE Joint Genome Institute.
 AUTHORS Direct Submission
 JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 3 (bases 1 to 123224)
 TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
 AUTHORS Direct Submission
 JOURNAL Submitted (20-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 On Jul 20, 2001 this sequence version replaced gi:13677003.
 Draft Sequence Produced by DOE Joint Genome Institute
 COMMENT www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center
 www.shgc.stanford.edu
 Quality: Phrap Quality >=40 99.7% of Sequence;
 Estimated Total Number of Errors is 0.3.
 STS Content:
 WI-14133 G22788.
 Location/Qualifiers
 1. 123224
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="CTB-83D3"
 BASE COUNT 43012 a 23877 c 23705 g 32630 t
 ORIGIN

Query Match 1.6%; Score 21; DB 9; Length 123224;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 682 AAAACGAGAAATGAGTTTC 702
 |||||
 Db 109559 AAAACGAGAAATGAGTTTC 109539

RESULT 42
 AC008491/c 143661 bp DNA PRI 31-MAY-2001
 LOCUS Homo sapiens chromosome 5 clone CTC-426L22, complete sequence.
 AC008491
 AC008491.6 GI:14269675
 VERSION
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 143661)
 TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
 AUTHORS Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 143661)
 TITLE DOE Joint Genome Institute.
 AUTHORS Direct Submission
 JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 3 (bases 1 to 143661)
 TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
 AUTHORS Direct Submission
 JOURNAL Submitted (31-MAY-2001) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 On May 31, 2001 this sequence version replaced gi:13676952.
 Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center
 www.shgc.stanford.edu
 Quality: Phrap Quality >=40 99.6% of Sequence;
 Estimated Total Number of Errors is 0.3.
 Location/Qualifiers
 1. 143661
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="CTC-426L22"
 BASE COUNT 45816 a 29494 c 29282 g 39069 t
 ORIGIN

Query Match 1.6%; Score 21; DB 9; Length 143661;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 682 AAAACGAGAAATGAGTTTC 702
 |||||
 Db 32913 AAAACGAGAAATGAGTTTC 32893

RESULT 43
 HS179115A 146810 bp DNA PRI 22-NOV-1999
 LOCUS Human DNA sequence from PAC 179115, BRCA2 gene region chromosome.
 DEFINITION 13612-913 contains Kiotho ESTs and Cpg island.
 ACCESSION 292540.1
 VERSION 292540.1 GI:2370068
 KEYWORDS 13612-q13; Cpg island.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 146810)
 TITLE McMurray/A. and Odell/C.
 JOURNAL Direct Submission
 Submitted (20-AUG-1997) Chromosome 13 Project Group
 (http://www.sanger.ac.uk/HGP/Chr13/) Sanger Centre, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk
 On May 14, 2001 this sequence version replaced gi:1869765
 gi:1869764.
 COMMENT IMPORTANT: This sequence is not the entire insert of clone 179115.
 It may be shorter because we arrange for a small overlap between
 once, or longer because we arrange for a small overlap between
 neighbouring submissions.
 This sequence was generated from part of bacterial clone contigs of
 human chromosome 13, constructed by the Sanger Centre chromosome 13
 mapping group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr13/
 This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.
 The true left end of clone 179115 is at 1 in this sequence. The
 true right end of clone 45010 is at 54668.
 The true left end of clone 46123 is at 146707.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variations annotated may not be found in the sequence submission
 corresponding to the overlapping clone as we submit sequences with
 only a small overlap as described above.
 179115 is from the library RPL1 constructed at the Roswell Park
 Cancer Institute by the group of Pictet de Jong.
 For further details see http://bacpac.med.buffalo.edu/.

FEATURES
 source 1. 146810
 Location/Qualifiers

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q12-q13"
/clone="XX-179115"
/clone_lib="RPC11"
1029..1334
/note="AluJo repeat: matches 296..2 of consensus"
repeat_region
1778..1945
/note="3 copies of 56 mer 85 & conserved"
repeat_region
3350..3496
/note="AluX repeat: matches 302..156 of consensus
incomplete repeat"
repeat_region
3606..3772
/note="MLT2_internal repeat: matches 4296..4125 of
consensus"
repeat_region
3808..4594
/note="MLT2_internal repeat: matches 4041..3339 of
consensus"
repeat_region
4627..5080
/note="MLT2_internal repeat: matches 3146..2693 of
consensus"
repeat_region
5238..5700
/note="MLT2_internal repeat: matches 2425..1983 of
consensus"
repeat_region
5727..5897
/note="MER42B repeat: matches 1109..1300 of consensus"
repeat_region
6013..6042
/note="15 copies of 2 mer 93 & conserved"
repeat_region
7140..7255
/note="49J10 data only"
repeat_region
8271..8292
/note="11 copies of 2 mer 100 & conserved"
repeat_region
9168..9279
/note="2 copies of 56 mer 84 & conserved"
repeat_region
9310..9576
/note="AluJo repeat: matches 257..1 of consensus
incomplete repeat"
repeat_region
9910..9949
/note="10 copies of 4 mer 85 & conserved"
repeat_region
10283..10356
/note="MLT1C repeat: matches 1..86 of consensus"
repeat_region
10354..11007
/note="LIMA6 repeat: matches 403..1042 of consensus"
repeat_region
11022..11165
/note="MLT1C repeat: matches 67..203 of consensus"
repeat_region
11022..11364
/note="MLT1B repeat: matches 67..390 of consensus"
repeat_region
11681..11793
/note="49J10 data only"
repeat_region
11924..12175
/note="MIR repeat: matches 259..11 of consensus"
repeat_region
12395..12604
/note="MIR repeat: matches 22..218 of consensus"
repeat_region
13477..13774
/note="AluX repeat: matches 1..298 of consensus"
repeat_region
13879..14366
/note="MLT1D repeat: matches 500..2 of consensus"
repeat_region
15236..15593
/note="MLT1A repeat: matches 365..1 of consensus"
repeat_region
16081..16179
/note="MIR repeat: matches 133..35 of consensus"
repeat_region
17340..17639
/note="AluJo repeat: matches 302..1 of consensus"
repeat_region
20484..20783
/note="AluX repeat: matches 302..1 of consensus"
repeat_region
21616..21916
/note="AluX repeat: matches 1..301 of consensus"
repeat_region
22923..23227
/note="MLT1G repeat: matches 512..235 of consensus"
repeat_region
22923..23026
/note="MLT1F repeat: matches 541..439 of consensus"
repeat_region
23230..23358
/note="MLT1G repeat: matches 160..31 of consensus"
repeat_region
25654..25859
/note="LIME3A repeat: matches 585..366 of consensus"
repeat_region
26810..26857
/note="24 copies of 2 mer 81 & conserved"
repeat_region
26939..27058
/note="AluJo repeat: matches 130..1 of consensus
incomplete repeat"
repeat_region
28240..28439
/note="MIR repeat: matches 223..20 of consensus"
repeat_region
28600..28699
/note="MIR repeat: matches 261..159 of consensus"
repeat_region
28700..29065
/note="MLT1A2 repeat: matches 374..1 of consensus"
repeat_region
29067..29208
/note="MIR repeat: matches 162..15 of consensus"
repeat_region
30635..30919
/note="MER34 repeat: matches 532..235 of consensus"
repeat_region
30925..31199
/note="MER34 repeat: matches 267..8 of consensus"
repeat_region
31050..31118
/note="49J10 data only"
repeat_region
31268..31344
/note="MIR repeat: matches 70..146 of consensus"
repeat_region
32875..33352
/note="49J10 data only"
repeat_region
34039..34333
/note="AluJo repeat: matches 293..1 of consensus"
repeat_region
37035..37341
/note="AluJo repeat: matches 5..302 of consensus"
repeat_region
38430..38586
/note="MIR repeat: matches 88..262 of consensus"
repeat_region
39109..39270
/note="MIR repeat: matches 47..207 of consensus"
repeat_region
39384..39740
/note="LIME1 repeat: matches 915..540 of consensus"
repeat_region
39536..40231
/note="LIME2 repeat: matches 744..2 of consensus"
repeat_region
40086..40364
/note="L1 repeat: matches 5390..5107 of consensus"
repeat_region
40435..40728
/note="AluJo repeat: matches 1..302 of consensus"
repeat_region
40770..41372
/note="L1 repeat: matches 5013..4394 of consensus"
repeat_region
41439..41728
/note="AluJo repeat: matches 296..1 of consensus"
repeat_region
41740..42353
/note="L1 repeat: matches 4313..3694 of consensus"
repeat_region
42420..42917
/note="MLT1D repeat: matches 1..503 of consensus"
repeat_region
44193..45080
/note="L1 repeat: matches 893..1 of consensus"
repeat_region
44935..45858
/note="L1 repeat: matches 5390..4467 of consensus"
repeat_region
44988..45003
/note="49J10 data only"
repeat_region
48909..48964
/note="MIR repeat: matches 144..88 of consensus"
repeat_region
48968..49571
/note="L1 repeat: matches 901..269 of consensus"
repeat_region
49815..49871
/note="7SLRNA repeat: matches 258..314 of consensus"
repeat_region
50360..50553
/note="MIR repeat: matches 18..228 of consensus"
repeat_region
52978..53072
/note="MER5A repeat: matches 103..7 of consensus"
repeat_region
53125..53426
/note="AluX repeat: matches 1..300 of consensus"
repeat_region
53734..53853
/note="MIR repeat: matches 25..146 of consensus"
repeat_region
54190..54318
/note="49J10 data only"
repeat_region
55591..56086
```

repeat_region /note="L1 repeat: matches 4844: .5349 of consensus"
55984.56140
/note="L1MB5 repeat: matches 5.157 of consensus"
56141.56434
repeat_region /note="Alusq repeat: matches 1.295 of consensus"
56436.57216
repeat_region /note="L1MB4 repeat: matches 148.932 of consensus"

Query Match 1.6%; Score 21; DB 9; Length 146810;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1175 AACGAGTGTGATTCGAGCT 1195
|||||
DB 73954 AACGAGTGTGATTCGAGCT 73974

RESULT 44
AC068315 155313 bp DNA HTG 27-SEP-2000
LOCUS Homo sapiens chromosome 3 clone RP11-224P21 map 3p, WORKING DRAFT
DEFINITION AC068315.4 GI:10312236
ACCESSION AC068315
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT;
KEYWORDS SOURCE
ORGANISM human
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 155313)
AUTHORS Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,
Dong, W., Fan, H., Feng, X., Guan, Q., Gu, X., Guo, D., He, L., Hu, S.,
Huang, F., Jin, Y., Kang, N., Li, C., Li, G., Li, J., Li, L.,
Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, Y.,
Luo, J., Niu, Y., Qi, Q., Qi, X., Song, S., Sun, M., Sun, W., Sun, Y.,
Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, R.,
Yan, C., Yang, X., Yu, B., Zeng, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y.,
Zhang, L., Zhang, M., Zhang, X., Zhang, G., Zhang, H., Zhang, Y.,
Zhang, Z., Zhu, B., Yu, J. and Yang, H.
Chromosome 3p genomic sequence
Unpublished
2 (bases 1 to 155313)
AUTHORS Bao, W., Hu, S., Dong, W., Wang, J., Zhang, Y., Zhang, H., Liu, B.,
Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y., Niu, Y.,
Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H., Liu, Y.,
Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L., Guo, D.,
Huang, F., Zhang, G., Li, T., Bian, X., Zhang, M., Li, L., Feng, X., Yu, J.
and Yang, H.
Direct Submission
Submitted (02-MAY-2000) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P.R.China
on Sep 27, 2000 this sequence version replaced gi:1011156.
Genome Center
Center: Beijing Center
Center code: Beijing
Website: http://hgsc.igbp.ac.cn
http://www.genomics.org.cn
Contact: hgsc@igbp.ac.cn
Project Information
Center project name: RP11-224P21
Center clone name: RP11-224P21
Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator; ET 5% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 154843 bases at least Q40
Consensus quality: 165011 bases at least Q30
Consensus quality: 168985 bases at least Q20
Insert size: 139544; sum-of-contigs
Quality coverage: 5.36x in Q20 bases; sum-of-contigs

***** NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available, and the accession number will
* be preserved.

1 1139: contig of 1139 bp in length
1140 1239: gap of unknown length
1240 2448: contig of 1209 bp in length
2449 2548: gap of unknown length
2549 3935: contig of 1387 bp in length
3936 4035: gap of unknown length
4036 6131: contig of 2096 bp in length
6132 6231: gap of unknown length
6232 7429: contig of 1198 bp in length
7430 7528: gap of unknown length
7529 9034: contig of 1525 bp in length
9035 9154: gap of unknown length
9155 10950: contig of 1796 bp in length
10951 11050: gap of unknown length
11051 12637: contig of 1587 bp in length
12638 12737: gap of unknown length
12738 15121: contig of 2384 bp in length
15122 15221: gap of unknown length
15222 17320: contig of 1999 bp in length
17321 20436: contig of 3116 bp in length
20437 20536: gap of unknown length
20537 22761: contig of 3225 bp in length
22762 23861: gap of unknown length
23862 26190: contig of 2329 bp in length
26191 26290: gap of unknown length
26291 28871: contig of 2581 bp in length
28872 28971: gap of unknown length
28972 33062: contig of 4091 bp in length
33063 33162: gap of unknown length
33163 37033: contig of 3871 bp in length
37034 37133: gap of unknown length
37134 43465: contig of 6332 bp in length
43466 43565: gap of unknown length
43566 48805: contig of 5240 bp in length
48806 48906: gap of unknown length
48907 56158: contig of 7253 bp in length
56159 63409: gap of unknown length
63410 63509: contig of 7151 bp in length
63510 72539: gap of unknown length
72540 72639: contig of 9030 bp in length
72640 83389: gap of unknown length
83390 83489: contig of 10750 bp in length
83490 91426: gap of unknown length
91427 91526: contig of 7937 bp in length
91527 99509: gap of unknown length
99510 99609: contig of 7983 bp in length
99610 108238: gap of unknown length
108239 108338: contig of 8629 bp in length
108339 119079: gap of unknown length
119080 119179: contig of 10741 bp in length
119180 135925: gap of unknown length
135926 136025: contig of 16746 bp in length
136026 155313: gap of unknown length
155313 155313: contig of 19286 bp in length.

FEATURES

source

1..155313 Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p"
/clone="RP11-224P21"
1..1139
/note="assembly_name:Contig32"

misc_feature 1240..2448
/note="assembly_name:Contig33"
misc_feature 2549..3935
/note="assembly_name:Contig34"
misc_feature 4036..6131
/note="assembly_name:Contig35"
misc_feature 6232..7429
/note="assembly_name:Contig36"
misc_feature 7530..9054
/note="assembly_name:Contig37"
misc_feature 9155..10950
/note="assembly_name:Contig38"
misc_feature 11051..12637
/note="assembly_name:Contig39"
misc_feature 12738..15121
/note="assembly_name:Contig40"
misc_feature 15222..17220
/note="assembly_name:Contig41"
misc_feature 17321..20436
/note="assembly_name:Contig42"
misc_feature 20537..23761
/note="assembly_name:Contig43"
misc_feature 23862..26190
/note="assembly_name:Contig44"
misc_feature 26291..28871
/note="assembly_name:Contig45"
misc_feature 28972..33062
/note="assembly_name:Contig46"
misc_feature 33163..37033
/note="assembly_name:Contig47"
misc_feature 37134..43465
/note="assembly_name:Contig48"
misc_feature 43566..48805
/note="assembly_name:Contig49"
misc_feature 48906..56158
/note="assembly_name:Contig50"
misc_feature 56259..63409
/note="assembly_name:Contig51"
misc_feature 63510..72539
/note="assembly_name:Contig52"
vector_end:SP6
vector_side:right"
misc_feature 72640..83389
/note="assembly_name:Contig53"
misc_feature 83490..91426
/note="assembly_name:Contig54"
misc_feature 91527..99509
/note="assembly_name:Contig55"
misc_feature 99610..108238
/note="assembly_name:Contig56"
misc_feature 108339..119079
/note="assembly_name:Contig57"
misc_feature 119180..135925
/note="assembly_name:Contig58"
misc_feature 136026..155313
/note="assembly_name:Contig59"
BASE COUNT 37836 a 38371 c 38396 g 38005 t 2705 others
ORIGIN

Query Match 1.6% Score 21 DB 2 Length 155313;
Best Local Similarity 100.0% Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 56 CTGTGCGAGCGCGCGCGAG 76
|||||
Db 127763 CTGTGCGAGCGCGCGCGAG 127783
RESULT 45
AC066599 157007 bp DNA HTG 19-SEP-2000
LOCUS Homo sapiens chromosome 3 clone RP11-596J9 map 3p, WORKING DRAFT
DEFINITION

SEQUENCE, 14 unordered pieces.
AC066599
AC066599.3 GI:10190763
KEYWORDS
HTG, HTGS_PHASE1, HTGS_DRAFT.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 157007)
REFERENCE
AUTHORS
Bao J., Bao Q., Bao W., Bian X., Cao T., Chen C., Chen J., Ding H.,
Dong W., Fan H., Feng X., Guan Q., Gu X., Guo D., He L., Hu S.,
Huang F., Jin Y., Kang N., Li C., Li G., Li J., Li L.,
Li S., Li T., Liu Y., Liu N., Liu B., Liu Y., Li W., Li Y.,
Luo J., Niu Y., Qi Q., Qi X., Song S., Sun M., Sun W., Sun Y.,
Tao R., Wang H., Wang J., Wang J., Wang L., Wang R.,
Wang X., Wang X., Wang Y., Wu D., Wu Q., Xie F., Xuan Z., Xue Y.,
Yan C., Yang X., Yu B., Zeng Y., Zhang G., Zhang H., Zhang H.,
Zhang L., Zhang M., Zhang X., Zhang Y., Zhang Y.,
Zhang Z., Zhu B., Yu J. and Yang H.
Chromosome 3p genomic sequence
Unpublished
2 (bases 1 to 157007)
Wang J., Hu S., Dong W., Wang J., Zhang Y., Zhang H., Liu B.,
Bao W., Sun Y., Wu Q., Wang H., Yang X., Cheng C., Wang Y., Niu Y.,
Qi X., Li T., Zhang H., Liu N., Wu D., Yu B., Fan H., Guo D.,
Li G., Li C., Bao Q., Bao J., Wang X., Song L., Zhang L., Guo D.,
Huang F., Zhang G., Li J., Bian X., Zhang M., Li L., Feng X., Yu J.
and Yang H.
Direct Submission
Submitted (25-APR-2000) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P.R.China
On Sep 19, 2000 this sequence version replaced gi:8101249.
COMMENT
-----Genome Center
Center:Beijing Center
Center code:Beijing
Website:http://hgsc.igtp.ac.cn
http://www.genomics.org.cn
Contact:hgsc@igtp.ac.cn
-----Project Information
Center project name:RP11-596J9
Center clone name:RP11-596J9
-----Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator Big Dye; 45% of reads
Chemistry: Dye-terminator Big Dye; 45% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 150014 bases at least Q40
Consensus quality: 153515 bases at least Q30
Consensus quality: 155718 bases at least Q20
Insert size: 148003; sum-of-contigs
Quality coverage: 4.98x in Q20 bases;sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1641: contig of 1641 bp in length
* 1642 1741: gap of unknown length
* 1742 5811: contig of 4070 bp in length
* 5812 5911: gap of unknown length
* 5912 7862: contig of 1951 bp in length
* 7863 7962: gap of unknown length
* 7963 11614: contig of 3652 bp in length
* 11615 11714: gap of unknown length
* 11715 15628: contig of 3914 bp in length
* 15629 15728: gap of unknown length
* 15729 22199: contig of 6471 bp in length
* 22200 22299: gap of unknown length

```
* 22300 32786: contig of 10487 bp in length
* 32787 32886: gap of unknown length
* 32887 39504: contig of 6618 bp in length
* 39505 39605: gap of unknown length
* 39605 48115: contig of 8511 bp in length
* 48115 48216: gap of unknown length
* 48216 57788: contig of 9573 bp in length
* 57788 57889: gap of unknown length
* 57889 69171: contig of 11283 bp in length
* 69171 69272: gap of unknown length
* 69272 86305: contig of 17034 bp in length
* 86305 86406: gap of unknown length
* 86406 117887: contig of 31482 bp in length
* 117887 117988: gap of unknown length
* 117988 157007: contig of 39020 bp in length.
```

FEATURES

Source

```
1. 157007
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="3"
  /map="3p"
  /clone="RP11:596J9"
```

```
misc_feature
  1. 1641
  /note="assembly_name:Contig5"
  1742. 5811
  /note="assembly_name:Contig6"
  5912. 7862
  /note="assembly_name:Contig7"
  7963. 11614
  /note="assembly_name:Contig8"
  11715. 15628
  /note="assembly_name:Contig9"
  15729. 22199
  /note="assembly_name:Contig10"
  22300. 32786
  /note="assembly_name:Contig11"
  32887. 39504
  /note="assembly_name:Contig12"
  39605. 48115
  /note="assembly_name:Contig13"
  48216. 57788
  /note="assembly_name:Contig14"
  57889. 69171
  /note="assembly_name:Contig15"
  69272. 86305
  /note="assembly_name:Contig16"
  86406. 117887
  /note="assembly_name:Contig17"
  117988. 157007
  /note="assembly_name:Contig18"
```

```
BASE COUNT 39725 a 38767 c 38530 g 38682 t 1303 others
ORIGIN
```

```
Query Match 1.6% Score 21; DB 2; length 157007;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 56 CTGTGGCAGCGCGCGCGCAG 76
Db 82733 CTGTGGCAGCGCGCGCGCAG 82753
```

```
Search completed: March 29, 2002, 21:43:28
Job time: 6623 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2002, 20:20:40 ; Search time 171.98 Seconds

(Without alignments)
6355.912 Million cell updates/sec

Title: US-09-415-540-2

Perfect score: 1275
Sequence: 1 CAACAGCTTNGGGCTCTCT.....GAAATAAATATTTGCTG 1275

Scoring table: OLIGO_NUC

Gapop 60.0, Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

N_Geneseq_1101.*
1: /SIDSI/gcgdata/geneseq/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1275	100.0	1275	20	AAZ23249
2	1275	100.0	1275	20	AAV63542
3	1275	100.0	1277	22	AA508596
4	1214	95.2	1351	20	AA500685
5	1153	90.4	1350	20	AA500716
6	1055	82.7	1200	20	AA533425
7	788	61.8	856	20	AA77501
8	659	51.7	660	21	AA10628
9	415	32.5	578	21	AA78135
10	415	32.5	578	22	AA128873
11	317	24.9	416	16	AA721836

12	202	15.8	357	20	AAZ77472	Human ovarian tumor
13	121	9.5	121	21	AAAC14929	Human secreted protein
14	68	5.3	102	21	AAAC4576	Human secreted protein
15	66	5.2	185	22	AA122766	Probe #12699 for g
16	66	5.2	185	22	AA148068	Probe #16754 used
17	66	5.2	185	22	AA108439	Probe #8430 used
18	66	5.2	411	22	AA113542	Probe #3475 for g
19	66	5.2	411	22	AA134904	Probe #3590 used
20	66	5.2	411	22	AA103431	Probe #3422 used
21	42	3.3	342	22	AAH35811	Human colon cancer
22	19	1.5	1198	20	AAH89287	Human regulatory p
23	19	1.5	1202	20	AAZ52867	Human prostate tum
24	19	1.5	1215	21	AAAC98069	Human colon cancer
25	19	1.5	1536	22	AA159678	Human polynucleoti
26	19	1.5	1536	22	AA159679	Human polynucleoti
27	19	1.5	2264	20	AAV81267	Bcl-2 interaction
28	19	1.5	2528	21	AAZ51797	Full length cDNA s
29	19	1.5	2534	21	AAZ51804	Full length expand
30	19	1.5	3636	20	AAV84234	Rat (Na,K)-ATPase
31	19	1.5	3636	20	AAV84235	Rat (Na,K)-ATPase
32	19	1.5	3636	20	AAV84236	Rat (Na,K)-ATPase
33	19	1.5	3636	20	AAV84237	Rat (Na,K)-ATPase
34	19	1.5	7328	22	AAZ44683	Novel protein kina
35	18	1.4	30	21	AAZ38953	Expression vector
36	18	1.4	346	20	AAV90298	EST clone D1482.
37	18	1.4	402	21	AAAC0150	Human secreted pro
38	18	1.4	543	21	AAE11454	Aspergillus niger
39	18	1.4	582	22	AAH13155	Human cDNA clone
40	18	1.4	588	22	AAH67523	C glutamicum codin
41	18	1.4	593	21	AAE11560	Aspergillus niger
42	18	1.4	624	17	AAE17915	VEGF-B16 coding s
43	18	1.4	624	20	AAV63569	cDNA clone encodin
44	18	1.4	624	22	AAAO1005	Human VEGF-3 codin
45	18	1.4	762	21	AAH68125	Pinus radiata flav

ALIGNMENTS

RESULT 1	
AAZ23249	AAZ23249 standard; cDNA; 1275 BP.
XX	XX
AC	AAZ23249;
XX	XX
DE	24-JAN-2000 (first entry)
XX	XX
DE	Human pyrophosphatase (HPP) encoding cDNA.
XX	XX
KW	Human; pyrophosphatase; HPP; genetic engineering; inflammatory disease;
KW	cancer; diagnostic assay; ss.
XX	XX
OS	Homo sapiens.
XX	XX
PN	US5981232-A.
XX	XX
PD	09-NOV-1999.
XX	XX
FE	13-AUG-1998; 9805-0134593.
XX	XX
PR	31-OCT-1996; 9605-0741437.
XX	XX
PA	(INCT-) INCTE PHARM INC.
XX	XX
PI	Hallman JL, Hawkins PR;
XX	XX
DR	WPI: 1999-633321/54.
XX	XX
DR	P-PSDB; AAY41160.
XX	XX
PT	Human pyrophosphatase used in treatment and prevention of cancer
XX	XX
PS	Disclosure; Fig 1A-D; 30pp; English.
XX	XX

This cDNA encodes a human pyrophosphatase (HPYP). HPYP is produced using genetically engineered expression vectors and host cells comprising the nucleic acid sequences encoding HPYP. HPYP and agonists, antibodies, or antagonists specifically binding HPYP are used in the prevention and treatment of cancer and inflammatory diseases. Antisense molecules to polynucleotides encoding HPYP are used for the treatment of diseases associated with the expression of HPYP. Polynucleotide, fragments, or the complement, and antibodies specifically binding HPYP can be used in diagnostic assays.

Sequence 1275 BP; 394 A; 235 C; 294 G; 351 T; 1 other:

Query Match 100.0%; Score 1275; DB 20; Length 1275;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 CAGAGGTTNGGGGCTCTCTCTCTGTCAGTCGGCCCGCGTGCGGCTGCTCTGTG 60
    |||
Db 1 caagaggttnggggctctctctctctctcagtcgagcgcgctggtggtctctgtg 60
OY 61 GGAGGGGGGGGGGCGGAGCTCCGACATATGAGGGGCTTCAGCAGCAGGAGCGCGCGC 120
    |||
Db 61 ggaagcgggggggggcgagagctccgacatagcggtctcagcagcgagcgcgcg 120
OY 121 GCCCTTCCCTCGAGATCCAGATCTTCTCAAAAATGAGAAAGACATATATCTCC 180
    |||
Db 121 gcccttccctcgagagctccgagctctcctcaaaaatgagaaagacatatactcc 180
OY 181 ATTTCATGATATTCATTTATTCAGATATGAGATGTGTTTCAGATGATGAAATAC 240
    |||
Db 181 atttcgatatttcgaattatgacataagatgtgtttcacaatggttgaagatcc 240
OY 241 AGCGTGTCTATGCAAAATGAGATGCTCTCAAAAGACCTTAAACCTTATTAACA 300
    |||
Db 241 agcggtgtctatgcaaaaatgagatgctctcaaaagacctttaaaccattaaaca 300
OY 301 AGATGTGAAAAAGAAAACTTCGTATGTTGCGAATTTGTTCCGTATTAAGATATAT 360
    |||
Db 301 agatgtgaaaaaagaaaaacttcgtatgttgcgaatttgcgtatataagatatat 360
OY 361 CTGGAACCTTGTGTCACCTCCCTGACCTTGGAGAACCCAGGGGCAATGATTAACATAC 420
    |||
Db 361 ctggaaccttgtgtcacctccctgaccttggagaagccagggcaaatgataaacatc 420
OY 421 TGCGTGTGTGATGACCAATGACCAATGATGTTGTAATGGAAGCAAGATATGTC 480
    |||
Db 421 tgcggtgtgtgatgacccaatgacccaatgattgtgtgaatggaaagcaagatgtgc 480
OY 481 AAGAGGTGAATATATGCGGTGAAAGTTCTAGGCAATTTGGCTATGATGACGAAGGGA 540
    |||
Db 481 aagaggtgaatatattggtgctgaaagtcttagacatattgtgacgaagggga 540
OY 541 AACCGCTGGAAAGTATTCGCAATTAATGATGATGATGATGATGATGATGATGATG 600
    |||
Db 541 aaccgctggaaagtatttcgcaattaatgatatgatatgatatgatatgatatg 600
OY 601 TATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
    |||
Db 601 tatcatgatgatgatgatgatgatgatgatgatgatgatgatgatgatgatgatg 660
OY 661 AAGGTATGAGTTCCGATGAGAAACAGAAATGAGTTGCGTTTAAATGACAAATTTAA 720
    |||
Db 661 aaggtatgagttccgatgagaaacagaaatgagtttgcgttttaatgcagaattaa 720
OY 721 AGATTAAGACTTTGCCATTTGATTTTAAAGACACTGACATTTGGAAGCAATTAAT 780
    |||
Db 721 agattaagactttgccatttgaatttttaaaagacctgacatttggaaagcaattaa 780
OY 781 GACTTAAGAAAGCAATGAGAAAGCAATGATGATGATGATGATGATGATGATGATG 840
    |||
Db 781 gacttaagaaagcaatgagaaagcaatgattgattgattgattgattgattgattg 840

```

```

OY 841 CTTCAGTGTGATTCGATGCTGCCAGAGCCATTGTGATGCTTTACACACCCTGTGCA 900
    |||
Db 841 cttaagtgatgctgattgctgagagccattgtgattgattgattgattgattgattg 900
OY 901 ATTCGCTGACAGCTTACCAACAGCAGTGTGATGCTTCCATCCACAGAAAGCTAATG 960
    |||
Db 901 attcgctgacagcttaccacagcgatgattgattgattgattgattgattgattg 960
OY 961 AGATTCTCTGGAATATACAGCTGATTTGCTACATCTGTTCACTTGTGATGATTAAG 1020
    |||
Db 961 agatttctctggaaatatacagctgatttgcattgctacatgctgattgattgattg 1020
OY 1021 TAAAGATGATGCTTTTCAAGCTTTAAATTTGAGAACATCTTACATGAAATTC 1080
    |||
Db 1021 taaagatgattgcttttcaagcttttaatttgaagaaatcttacaatgaaatct 1080
OY 1081 TGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140
    |||
Db 1081 tgctgtgatgattgattgattgattgattgattgattgattgattgattgattgattg 1140
OY 1141 ACTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200
    |||
Db 1141 actaagatgattgattgattgattgattgattgattgattgattgattgattgattg 1200
OY 1201 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260
    |||
Db 1201 gtgatgattgattgattgattgattgattgattgattgattgattgattgattgattg 1260
OY 1261 AAAATATTTTGTGCTG 1275
    |||
Db 1261 aaaattatttgcgtg 1275

```

RESULT 2
ID AAV63542 standard; cDNA; 1275 BP.
AC AAV63542;
DT 03-FEB-1999 (first entry)
DE cDNA encoding a human inorganic pyrophosphatase designated HPYP.
KW Human; inorganic pyrophosphatase; HPYP; Incyte clone 768320;
KM cell regeneration; tissue regeneration; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 89..958 /tag= a
FT /product= HPYP
PN US843665-A.
PD 01-DEC-1998.
PE 31-OCT-1996; 96US-0741437.
PR 31-OCT-1996; 96US-0741437.
PA (INCY-) INCYTE PHARM INC.
PI Hawkins PR, Hillman JL.
DR MPI: 1999-044570/04.
DR P-PSDB: AAW70876.
PT DNA encoding human pyrophosphatase protein -- useful for producing
PT recombinant protein, which may be useful in promoting cell and
PT tissue regeneration
PS Claim 3; Fig 3A-D; 29pp; English.

XX The present sequence encodes a human inorganic pyrophosphatase
 CC designated HPYP. Nucleic acids encoding HPYP were first identified
 CC in Incyte clone 768320 from the lung tissue cDNA library LUNGNOT04.
 CC The recombinant pyrophosphatase, may be useful in promoting cell
 CC and tissue regeneration.

SO Sequence 1275 BP; 394 A; 235 C; 294 G; 351 T; 1 other;

Query Match 100.0%; Score 1275; DB 20; Length 1275;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGAGTTGGGAGCTCTCTTGTGAGTCGGCGCGGCTGGGGCTGGTCTGTG 60
 Db 1 caagagttgggagctctcttgtgagtcggcgcggtggggctgggtctgtg 60
 QY 61 GCAGCGGCGGCGAGAGACTCCGCACTATGACGCGCTTCAGCAGCAGAGCGCGC 120
 Db 61 gcagcgcgcgcgagagactccgcactatgagcggttcagcagcagagcgcgcg 120
 QY 121 GCCCTTCCTCCTGAGTACCGAGCTTCTCTCAAAATGAGAAAGACATATATCTCC 180
 Db 121 gcccttcctcctgagtagccgagctctctcaaaaatgagaaagacatatatctcc 180
 QY 181 ATTTCATGATTCGCAATTTATGCAATAGAGATGCTTTCACATGATTTGAATACC 240
 Db 181 atttcgatattcgcaatttatgcaatagagatgctttcacatgatttgaataacc 240
 QY 241 ACCGCTGCTAATGCAAAATGAGATTGTACAAAGACCCCTTAAACCTTTAAACA 300
 Db 241 accgctgctaatgcaaaatgagattgtacaaagacccctttaaaccctttaaca 300
 QY 301 AGATGTGAAAAAGAAAACCTGCTATGTGGCAATTTGTCCTGATTAAGATATAT 360
 Db 301 agatgtgaaaaagaaaacctgctatgtggcaattttgtccctgataaagatatat 360
 QY 361 CTGGACTATGCTGCATCCCTGCACTTGGGAAGCCGACACCAATATTAACCTAC 420
 Db 361 ctggactatgctgcataccctgcacttgggaagcccgacacccaatatttaacctac 420
 QY 421 TGGCTGTGTGTGACAAATGACCAATGTGTGTGAAATTTGGAAGCAAGTATGTC 480
 Db 421 tggctgtgtgtgacaaatgaccatgtgtgtgaaattttggaagcaagtatgtgc 480
 QY 481 AAGAGTGAATTAATTTGGCGTGAAGTTCTAGCAGATTTGGCTATGATTAACGAG 540
 Db 481 aagagtgaaatatttggcgtgaagttctagcagatttggctatgattgacgaagg 540
 QY 541 AACCGACTGGAAGTCAATTCCTTAATGATGATGATCGACGCAATTAATGA 600
 Db 541 aaccgactggaaagtcaattccttaatgatgatgatcgacgcaatttaattga 600
 QY 601 TATCAATGATGTAACAGCGCTGAACCTGCTACTTGAAGACTACTGTGACTGTTAG 660
 Db 601 tatcaatgatgttaacagcgctgaacctgctactttgaagactactgtgactgttag 660
 QY 661 AAGGTATAGGTTCCATGATGAAAAACCAAAATGAGTTTGCTTTAATGACAAATTA 720
 Db 661 aaggtataggttccatgatgaaaaaccaaattgagtttgcttttaattgacaaatt 720
 QY 721 AGATAGGACTTGGCAATATATATTAATAACACATGACCATTTGGAAGCATTAAGT 780
 Db 721 agataggacttggcaattatattatataaagcaccatgacattggaagcattagt 780
 QY 781 GACTAAGAAAAGCAATGAGAAAGCAATGATGATGATGATGATGATGATGATGATG 840
 Db 781 gactaagaaaagcaatgagaaagcaatgattgattgattgattgattgattgattg 840
 QY 841 CTTCAAGTGTGATCTGATGCTGCTGACAGGACCATTTGATGATGATGATGATGATG 900
 Db 841 ctccaagtgtgattctgattgctgctgacagacccatttgattgattgattgattg 900

QY 901 ATCTGCTGCACTTACCAACAGAGCTGATAGTGGTTCATCCACAGAAAACTAATG 960
 Db 901 atctgctgcacttaccacagagctgatagtggttcattccacagaaaaactaatg 960
 QY 961 AGATTTCTGGAATFACAGCTGATATGCTACATCGTGTTCATCTGATGATTAAG 1020
 Db 961 agatttctggaaatfaccagctgatattgctacatcggtgttcattctgattagaag 1020
 QY 1021 TAAAGTAGTAGCTTTTCAAAAGCTTTAAATTTGAGAACTCATTAACCTAAAGTTC 1080
 Db 1021 taaagtagtagcttttcaaaagcttttaatttgagaaactcatctaagaaagtttc 1080
 QY 1081 TGGTGTGACTTAATTCATATACATGATGATGATGATGATGATGATGATGATGATG 1140
 Db 1081 tgggtgtgacttaatttcattatgattgattgattgattgattgattgattgattg 1140
 QY 1141 ACTAAGTAACTTTTACACATGCTTAAATATCAACAGCTTGTCAATTTGGAAGTCACT 1200
 Db 1141 actaagttaacttttaccatgctttaaattcaaacagcttgtcaatttggaaagtc 1200
 QY 1201 GTCAATGATGTGCAAGGAGACACATTTGATGATGATGATGATGATGATGATGAT 1260
 Db 1201 gtcaatgatgtgcaaggagacacatttggatgattgattgattgattgattgattg 1260
 QY 1261 AAAATTTTGTCTG 1275
 Db 1261 aaaatttgtctg 1275

RESULT 3
 AAS08596
 ID AAS08596 standard; cDNA; 1277 BP.
 AC AAS08596;
 XX 24-OCT-2001 (first entry)
 DE DNA encoding human pyrophosphatase (HPYP).
 KW Pyrophosphatase; human; HPYP; antiinflammatory; cytosolic;
 KW immunosuppressive; antianaemic; antidiabetic; osteopathic;
 KW antineumatic; antiathritis; inflammatory disease; cancer; nerve;
 KW transplantation; pancreatic; bone marrow; prostate; anemia; asthma;
 KW systemic lupus erythematosus; myasthenia gravis; diabetes mellitus;
 KW osteoporosis; glomerulonephritis; rheumatoid arthritis; osteoarthritis;
 KW scleroderma; ss.
 OS Homo sapiens.
 XX
 FH Key 1 Location/Qualifiers
 FT CDS 91..960
 FT /tag= a
 FT /product= "Human pyrophosphatase"
 US2001010911-A1.
 PD 02-AUG-2001.
 XX
 PF 08-OCT-1999; 99US-0415540.
 PR 31-OCT-1996; 96US-0741437.
 PR 13-AUG-1998; 98US-0134593.
 XX
 PA (HAWK) HAWKINS P R.
 PA (HILL) HILLMAN J L.
 PI Hawkins PR, Hillman JL;
 XX
 DR WPI; 2001-502139/55.
 DR P-PSDB; AA005122.
 XX
 PT New human pyrophosphatase polypeptides and polynucleotides, useful for

OY 1271 TCCTG 1275
11111
Db 1280 Tgctg 1284

RESULT 5

AXX00716
ID AAX00716 standard: DNA: 1350, BP.

AXX00716;

25-MAR-1999 (first entry)

Human secreted protein gene 75, clone H0SE145.

Human: secreted protein; fusion protein; gene therapy; protein therapy;
diagnosis: tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

Homo sapiens.

WO9842738-A1.

01-OCT-1998.

19-MAR-1998; 98WO-US05311.

30-MAY-1997; 97US-0050937.
21-MAR-1997; 97US-0041276.
21-MAR-1997; 97US-0041277.
21-MAR-1997; 97US-0041281.
21-MAR-1997; 97US-0042344.
30-MAY-1997; 97US-0048069.
30-MAY-1997; 97US-0048094.
30-MAY-1997; 97US-0048095.
30-MAY-1997; 97US-0048096.
30-MAY-1997; 97US-0048099.
30-MAY-1997; 97US-0048131.
30-MAY-1997; 97US-0048135.
30-MAY-1997; 97US-0048154.
30-MAY-1997; 97US-0048160.
30-MAY-1997; 97US-0048186.
30-MAY-1997; 97US-0048187.
30-MAY-1997; 97US-0048188.
30-MAY-1997; 97US-0048350.
30-MAY-1997; 97US-0048351.
30-MAY-1997; 97US-0048352.
30-MAY-1997; 97US-0048355.
05-AUG-1997; 97US-0054804.

(HUMA-) HUMAN GENOME SCI INC.

Brewer LA, Duan R, Ebner R, Ferrie AM, Florence KA;
Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS;
Rosen CA, Ruben SM, Shi Y, Young P;

WPI: 1999-070066/06.
P-PSDB: AAM67912.

New isolated human genes and the secreted polypeptides they encode -
useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders

Claim 1: Page 264-265; 385pp; English.

This sequence represents a nucleic acid molecule which encodes a
secreted human protein. The gene number, and the clone it is derived

from, are detailed in the descriptor line. The gene can be used to
generate fusion proteins by linking to the gene to a human immunoglobulin
Fc portion (e.g. AAX00602) for increasing the stability of the fused
protein as compared to the human protein only.
The invention relates to 87 novel genes and their fragments (nucleic
acid sequences: AAX00611-X00724; amino acid sequences AAM67807-N68004)
which are useful for preventing, treating or ameliorating medical
conditions e.g. by protein or gene therapy. Also, pathological
conditions can be diagnosed by determining the amount of the new
polypeptides in a sample or by determining the presence of mutations in
the new polynucleotides. Specific uses are described for each of the 87
polynucleotides, based on which tissues they are most highly expressed in
(see AAX00611 for described uses).

Sequence 1350 BP; 414 A; 250 C; 319 G; 363 T; 4 other;

Query Match 90.4%; Score 1153; DB 20; Length 1350;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 123 CCTTCTCCCTGAGTACCGAGTCTTCCCAAAATGGAAGACAAATATATCTCCAT 182
Db 131 ccttctccctgagtagcagagcttcccaaaatggaagacaaatatactccat 190
OY 183 TTCATGATATTCATTTATGACATTAAGATGTTGTTACATGTTAGTTAGTACCAC 242
Db 191 ttcatgatattccaatltatgacagtagagatggtttcacatgtagtgaagtagcac 250
OY 243 GCTGCTTAATGCAAAATGAGATGTTGTAAGAGACCCCTTAACCTTTAAACAG 302
Db 251 gctgcttaatgcaaaatgagatggttgaagagacccctttaaacccttaaacag 310
OY 303 ATGCAAAAAGAAACCTGCTATGCGAATTTGTTCCGATTAAGATATATCT 362
Db 311 atgcaaaaagaaaccttgcatagttggaattttccgataaagatatatct 370
OY 363 GGAAGTATGTCCTATCCCTCAGACTTGGGAAGACCCGAGCACATATATATCTG 422
Db 371 ggaactatgctgccaatccctcagacttgggaagaccgagacacatgataacactatg 430
OY 423 GCTGTTGTTGTCGACATGACCAATGATGTTGTAATGGAATGGAAGCATGTCGA 482
Db 431 gctgttggtggtgacatgacccaatgattggtgtaattgggaagcaagtagtgcaa 490
OY 483 GAGTGAAATTAATGCGGTGGAAGTTCTAGGCAATTTGCTATGATTAACGAAGGAAA 542
Db 491 gagtgaaataatttggtggaagttctagacataattgtagatgtagcgaagggaaa 550
OY 543 CCGACTGGAAGTCAATGCGATTATATGATGATGATGATGATGATGATGATGATG 602
Db 551 ccgactggaaagtcattgcattatgtagatgtagcgaagcgaatataatgata 610
OY 603 TCAATGATGTCGAACGGCTGGAACCTGGCTACTTGAAGCTACTGTGAGTGTAGAA 662
Db 611 tcaatgatgtcgaacggctggaacctggctacttgaagctactgtgagctggttagaa 670
OY 663 GGTATAAGTTCCTGATGGAAGAACAGAAATGAGTTGCTTAAATGCAATTTAAG 722
Db 671 ggtataagttcctgataggaagaaacagaaatgagtttgcattatgcaagaattaaag 730
OY 723 ATAGAGACTTTCCTGATGATTTAATAAGACATGATGATGATGATGATGATGATG 782
Db 731 ataagacttctgcatgatalatataaagacacatgacatggaagaagataagaga 790
OY 783 CTAAAGAAAGATGGAAGAAATCAGTTGATGATTAACAATTTGCTGAGAGCCCT 842
Db 791 cttaagaaacgatggaagaaatcagttgcatgatacaactgtgtcgtgagagccct 850
OY 843 TCAAGTGTATCTGATGCTGCTGCAAGCCATGTTGATGATGATGATGATGATGAT 902
Db 851 tcaagtgtatctgattgctgctgcaagccattgtgtagtattacacacacccctgtgaat 910

QY 903 CTGCTGACAGTACCAACAGACGTGATAAGTGTTCATCCACAGAAAACTAATGAG 962
D 911 CTGCTGACAGTACCAACAGACGTGATAAGTGTTCATCCACAGAAAACTAATGAG 970
QY 963 ATTTCTGTGAATACCAAGCTGATATTCCTACATCGTGTTCATCGATTAAGAAGTA 1022
D 971 attctctggaatacaagctgatatctacatcgtgtcactcgtatgtagaagta 1030
QY 1023 AAGTACTACTTTTCAAGCTTAAATTTGTAGACATCTACAAAGTAATCTG 1082
D 1031 aagatagagcttctcaagcttcaatctgtagaactcactcaacaaagtaattctg 1090
QY 1083 CTGTGATTAATCCAAATTAATCTACAGATTTATCCATCTAAAGCAATTTTTCATCTCAAC 1142
D 1091 ctggaactaactcaataactacagaaatgtaactcaactcaactcattcactcaac 1150
QY 1143 TAAGATACTTTTGGACATGCTTAATATCAAAAGCAAGTGTCTTTGGAAGTACTTGT 1202
D 1151 taagataacttctgacacatgcttaaatcaaaagcagctgtcattcgtgaagtaactgt 1210
QY 1203 GAATAGATGTGCAAGGAGACATATTTGATGTATGTATGTTACCATATGTTAGAAATTA 1262
D 1211 gaataagatgcaagggagacacatgtgatatgtatgtaccatattgttagaataa 1270
QY 1263 AATTATTTTCTGCTG 1275
D 1271 aattatttctg 1283

RESULT 6

AAZ33425 standard; cDNA: 1200 BP.

AAZ33425;

08-DEC-1999. (first entry)

Human prostate cancer-associated EST 3.

Expressed sequence tag: EST; prostate tumor; antitumor; treatment; gene therapy; tissue specificity human; ss.

Homo sapiens.

DE19811193-A1.

16-SEP-1999.

10-MAR-1998; 98DE-1011193.

10-MAR-1998; 98DE-1011193.

(META-) METAGEN GES GENOMFORSCHUNG MBH.

Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;

WPI: 1999-519628/44.

P-PSDB: AAY48218.

New nucleic acid expressed at high level in prostatic tumor tissue and encoded polypeptides, useful for treating cancer and screening for therapeutic agents

Claim 1a; 71: 16pp; German.

This invention describes novel nucleic acid sequences (A) that are expressed at high level in prostatic tumor tissue and encode gene products or their fragments. The products of the invention have antitumor activity. Polypeptides (I) encoded by (A) are used: (i) for identifying agents for treatment of prostatic cancer and (ii) for therapy of prostate cancer, optionally where expressed by gene therapy methods. (A) is also used to isolate full-length genes (for gene therapy) and for recombinant production of (I), which can be used to raise

CC specific antibodies. (A) are identified by assembly of ESTs (expressed sequence tags) before they are analyzed for expression pattern (tissue specificity). This approach eliminates many of the false results, as CC regards tissue specificity, associated with known methods that use CC single (usually short) ESTs. AAZ33423-533476 represent expressed CC sequence tags described in the method of the invention.

SQ Sequence 1200 BP; 368 A; 227 C; 278 G; 327 T; 0 other;

Query Match

82.7%; Score 1055; DB 20; Length 1200;

Best Local Similarity 100.0%; Pred. No. 0; Matches 1055; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CCTCTCCCTGAGTACCGAGTCTTCCCAAAATGAGAAAGCAATATATCTCAT 182
D 111 cctctccctgagtagacagagcttcccaaaaagaaagaaatataatctccat 170
QY 183 TTGATGATATTCATATTTATGCAATGAAGATGTTTACATGCTAGTGAATACG 242
D 171 ttcatgatattccaattatgcagataagatggtttcacatgtagtgaagtaacac 230
QY 243 GCTGCTTAATGCAAAATGAGATGCTTACAAAGACCTTTAAACCTTTAAACAG 302
D 231 gctgctcaatgcaaaaatgagatgcttacaagagacctttaaaccctataacaa 290
QY 303 ATGTGAAAAAGAAACCTTGCCTATGTTGGAATTTGTCCTGATTAAGATATCT 362
D 291 atgtgaaaaaagaaaccttgcctatgttggaattgttccgtataaagataatct 350
QY 363 GGAATATGTCGTCATCCCTCAGACTTGGAAAGCCAGGACCAATGATAACTACTG 422
D 351 ggaactatgctcattccctcagacttgggaagccagggacaatataacactctg 410
QY 423 GCTGTGTGTCGACAAATACCCATTTGATGTGGAATTTGGAAGCAAGTATGCA 482
D 411 gctgtgtgtgacaaatgcccaattgattgttgaatttgaagcaaggtatgtgcaa 470
QY 483 GAGGAAATATTTGGCTGAAAGTGTAGCATTTGCTATGATGATGACGAAGGGAAA 542
D 471 gaggtaaatatttggctgaaaggttcagacatatgtgtatgttgcgaagggaaa 530
QY 543 CCGACTGGAAGATCATTCCTATTAATGTGATGATCCATGATGACACCAATTATATG 602
D 531 ccgactggaagatcatcttccatattatgtgtatgtccgatgagccatataatgta 590
QY 603 TCAATGATGTCAAAGGCTGAACCTGGCTACTTGAAGTACTGTGAGCTGTTAGAA 662
D 591 tcaatgatgtcaaaagctgaacctgtacttgaagactatgtgacctgtttagaa 650
QY 663 GGTATTAAGCTTCGATGGAAGAACGAAATGAGTTGGCTTAATGACAGATTAAAG 722
D 651 ggtataagcttcgtagtgaagaaacagaaatgagtttgcgtttaaagcaaatltaag 710
QY 723 ATAAGGACTTTGCCATGATATTAATTAAGACACATGACCATTTGGAAGCATTAAG 782
D 711 ataagactttgccatgatattttaaaagacacatgacatgtgaagcatltagtga 770
QY 783 CTAAGAAAAGCAATGGAAGAAAGCATCACTGATGAATACAACTTTGCTGAGACCCCT 842
D 771 ctaagaaaagcaatggaagaaagcaatcagltgcatgaatacaacttgtctgagagccct 830
QY 843 TCAAGTGTGATCGATGCTGCTGCGAGAGCATTTGCGATGCTTACACACCTGTGAT 902
D 831 tcaagtgtgattcgtgagctgcccagagcatgtgtagtgccttaccaccctgtgat 890
QY 903 CTGCTGACAGTACCAACAGACGTGATTAAGTGTTCATCACACAGAAAACCTAATGAG 962
D 891 ctgctgacagtagtaccacagagctgtagtgaaggtttccatcaccaagaaaactatgag 950
QY 963 ATTTCTGTGAATACCAAGCTATATTTGCTACATCGTGTTCATCGATGATTAAGAAGTA 1022
D 951 attctctgatacaagctgatatgttaccatcgtgtcactcgtgtagtattagaagta 1010

[illegible]

OS Homo sapiens.
 XX EPI033401-A2.
 XX 06-SEP-2000.
 XX 21-FEB-2000; 2000EP-0200610.
 XX 26-FEB-1999; 99US-0122487.
 XX (GEST) GENSET.
 PA Dumas Milne Edwards J, Duclert A, Giordano J;
 XX WPI: 2000-500381/45.
 DR
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1; SEQ ID 14703; 71bp + CD-ROM; English.
 XX
 CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 CC
 SQ Sequence 660 BP; 218 A; 112 C; 133 G; 197 T; 0 other;
 Query Match 51.7%; Score 659; DB 21; Length 660;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 617 CGGGTGAACCTGGCTACTGATGAGTACCTGAGCTGTTAGAGGTATAGTCTCT 676
 Db 1 cggctgaacctgctacttaagaagctacgctgacgtggtcttaagaagctaaagcttct 60
 Oy 677 GATGGAACCCAGAAATGATGCTTGCCTTATGACAGATTAAAGATAGACTTGGC 736
 Db 61 gatggaacccagaagaatgagttgctgttaatgacagatttaagaagactttgccc 120
 Oy 737 ATTGATATTATTAAGACATCATGACCATTTGGAAGCATTAGTACTAGAAAAAGCAT 796
 Db 121 atgtatattataaagaacatcagcattggaagcattagtcactaagaagaacgat 180
 Oy 797 GGAAGAGAAATCATGATGATATACAACTTGTCTGAGACCCCTTAAAGTGTGATCT 856
 Db 181 ggaagaaagaaatcagttgatatacaacttgcitgagagcccttcaagtgtatcct 240
 Oy 857 GATGCTGCACAGCATTTGATGCTTACCAACCCCTGTGATCTGCTGCACATTA 916
 Db 241 gatctgcacagacatctgtgacttaccacaccccttgaatcgtcgtcacagta 300
 Oy 917 CCAACAGACGTGATTAAGTGGTTCATACCAAGAAAACTATGATGATTCTCTGGAATA 976
 Db 301 ccaacagacgtgataaagtggttccatcaccaagaataactatgacttctcgtgaata 360
 Oy 977 CAAGCTGATTTGCTACCTGCTGTTTCATCTGATGATATTAAGTAAGTAAGTACTT 1036
 Db 361 caagctgatcttctacactcgtgttcacatcgtatcttagaagaagaagtagtactt 420
 Oy 1037 TCAAGACTTTAAATTTGTAGAACATCATCTAACTAAGTAATCTGCTGCTACTATATCA 1096

Db 421 tcaaaagcttaatttgaagaacctactcaactaagaataatcgcgtgactaatcca 480
 Oy 1097 ATATACAGAAATGATTAACCATCTTAACCATTTTTCATATCTCACTAAGTAATCTTTA 1156
 Db 481 atatactagaatgttaccatcctaagaacttcttccatccactaagaataacttca 540
 Oy 1157 GCACATGCTTAAATATCAACAGCAGTTGTCAATTTGGAAGTCACTTGTGAATGATGCA 1216
 Db 541 gcacatgtcttaataataccaagaagcagttgcatcttggaaatcacttggatgatgtgcaa 600
 Oy 1217 GGGGAGCACAATTTGATGATGATATGTTACCATATCTTGAAGAAATAATATTTTGTG 1275
 Db 601 ggggagcacatattgataatgtatattgtaccatattgtagaataaatatttgcgtg 659
 RESULT 9
 AAA/8135/c
 ID AAA/8135 standard; cDNA: 578 BP.
 XX
 AC AAA/8135;
 XX
 DT 14-NOV-2000 (first entry)
 XX
 DE cDNA encoding human colon tumour polypeptide, SEQ ID NO:422.
 XX
 KW Human colon tumour polypeptide; tumour antigen; cancer; vaccine;
 KW immunotherapy; diagnosis; progression; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200037643-A2.
 XX
 PD 29-JUN-2000.
 XX
 PF 23-DEC-1999; 99WO-US30909.
 XX
 PR 23-DEC-1998; 98US-0221298.
 PR 02-JUN-1999; 99US-0347496.
 PR 22-SEP-1999; 99US-0401064.
 PR 19-NOV-1999; 99US-0444242.
 PR 02-DEC-1999; 99US-0454150.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk J;
 PI Wang T, Yuguin J;
 XX
 DR WPI: 2000-442671/38.
 XX
 PT New colon tumor polypeptides used to inhibit the development of cancer,
 PT especially colon cancer, and for diagnosing and monitoring the
 PT progression of the cancer -
 XX
 PS Claim 1; Page 212; 229pp; English.
 XX
 CC Sequences AAA77722-A78199 represent 478 cDNAs encoding proteins or
 CC portions of proteins which are associated with human colon tumours.
 CC The invention also specifically discloses 8 human colon tumour proteins
 CC (AAB11897-811904). The nucleic acids, the polypeptides they encode, and
 CC antigen presenting cells (APCs, preferably dendritic cells) expressing
 CC such polypeptides may be used in vaccines that target tumour cells,
 CC especially colon tumour cells, thereby inhibiting the development of
 CC cancer. T-cells specific for the polypeptide expressed by the APC are
 CC used to remove tumour cells from biological samples, especially blood or
 CC fractions thereof. The sample or the isolated T-cells specific for the
 CC polypeptide can then be used to inhibit cancer development. CD4+ and/or
 CC CD8+ T-cells from a patient may be incubated with a polypeptide or
 CC nucleic acid of the invention, or an APC expressing such a polypeptide,
 CC to cause the proliferation of specific T-cells. The T-cells can be
 CC cloned and then administered back to the patient to inhibit cancer
 CC development. Nucleic acids encoding the polypeptides and antibodies
 CC against the polypeptides may be used to determine the expression level

AT21836	AA21836 standard; cDNA to mRNA; 416 BP.
XX	AA21836:
AC	01-AUG-1996 (first entry)
DT	Human gene signature HOMGS03377.
XX	
DE	Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW	human; cloning; mapping; non-biased library; diagnosis; detection;
KW	cell typing; abnormal cell function; ss.
XX	
OS	Homo sapiens.
XX	
PN	W09514772-A1.
XX	
PD	01-JUN-1995.
XX	
PF	11-NOV-1994; 94WO-JP01916.
XX	
PR	12-NOV-1993; 93JP-0355504.
XX	
PA	(MATSU) MATSUBARA K.
PA	(OKUBO) OKUBO K.
XX	
PI	Matsubara K, Okubo K;
DR	WPI: 1995-206931/27.
XX	
PT	Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT	for diagnosis of abnormal cell function, by preparing cDNA that
PT	reflects relative abundance of corresp. mRNA in specific human
PT	tissues
XX	
PS	Claim 1; Page 977; 2245pp; Japanese.
XX	
CC	A single-stranded DNA (or its complementary strand or the corresp.
CC	double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC	given in AAT19001-126837 and which is able to hybridise to part of
CC	human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC	sequences were obtained from 3'-directed cDNA libraries prepared
CC	from various human tissues; synthesis of cDNA was initiated from the
CC	3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC	untranslated sequence is unique to a particular mRNA species, almost
CC	all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC	is constructed so as to reflect accurately the relative abundance of
CC	different mRNAs in the particular tissue from which it was derived.
CC	The appearance frequency of a given GS in a cDNA library can be
CC	determined (esp. using primers and probes derived from the GS
CC	sequences) as a means of diagnosing abnormal cell function or for
CC	recognising different cell types.
XX	
XX	
Q0	Sequence 416 BP; 128 A; 77 C; 77 G; 128 T; 6 other;
	Query Match
	Best Local Similarity 99.7%; Score 317; DB 16; Length 416;
	Matches 367; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	851 GATCCTGATGCTGCAGACCATTTGTGAGTGCCTTACACACACCTGTGAATTCGCTGC 910
Db	1 gatctgatctgtccagagccatltgtgatttaccacacacccctgtgacttgccttc 60
QY	911 ACAGTACCAACAGAGCTGTGATGATGAGTGTTCCTACCCAGAAAACCTAATGACATTTCTT 970
Db	61 acagtaaccaacagagctgtgataaagtggttccatccaccagaaaaactaaagatttctt 120
QY	971 GGAATACAAGCTGATATTGCTACATCGTTCATCTGATGATTTAGTAAGTAAAGTACT 1030
Db	121 ggaatacancgtcatattgtctacatccglttcatcttgatgtattgataagaagtaagt 180
QY	1031 AGCTTTCAAGACCTTTAAATTTGTAGAACCTCATCTACTAAGTAATTAATTTCTGCTGACT 1090

Query Match	9.5%	Score 121:	DB 21:	Length 121:
Best Local Similarity	100.0%	Pred. NO. 1.3e-50:		
Matches 121:	Conservative	0:	Mismatches 0:	Indels 0:
				Gaps 0:
QY 285	TAACACCTATTAAACAGATGTGAAAAAAGCAAAACCTTCGCTATGTGCGAATTTGTTC	344		
Db 1	taaacctatlaacaagaigtgaaaaaaggaaacttcgctatgttcgcaattgttcc	60		
QY 345	CGATTAAGGATATATCTGGACCTGTGTCGCATCCCTCAGACTTGGAGACCCAGGC	404		
Db 61	cgatlaaagatatatcttggaaactatgtgtccatccctcagaacttgggaagaccacgaagc	120		
QY 405	A 405			
	-			
Db 121	a 121			
RESULT 14				
AAAA5576				

AC	AAAA5576;
XX	
DT	21-AUG-2000 (first entry)
XX	
DE	Human secreted expressed sequence tag SEQ ID NO:2151.
XX	
XX	Human; mouse; chicken; rat; secreted expressed sequence tag; SESTR;
KM	expressed sequence tag; EST; probe; chemotactic; proliferative;
KM	immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KM	thrombolytic; antiinflammatory; cytosaric; antibacterial; antifungal;
KM	antiviral; antidiabetic; antisthmatic; vulnerrary; antiparkinsonian;
KM	antitumor; osteopathic; neuroprotective; nootropic; antipsoriatic;
KM	cardioprotective; anticonvulsant; antidepressant; gene therapy;
KM	vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KM	lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KM	central nervous system disorder; Alzheimer's disease; stroke;
KM	Parkinson's disease; Huntington's disease; coagulation disorder;
KM	haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW	tumour; infection; depression; psoriasis; ss.
OS	
XX	Homo sapiens.
XX	
PN	WO200021991-A1.
XX	
PD	20-APR-2000.
XX	
XX	15-OCT-1999; 99WO-US24206.
PF	
XX	15-OCT-1998; 98US-0104436.
PR	
XX	
XX	(GENM) GENETICS INST INC.
PA	
XX	Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI	Merberg D, Treacy M, Bowman MR;
PI	
XX	
DR	WPI: 2000-317938/27.
XX	
XX	Isolated polynucleotides, and encoded proteins, comprising secreted
PT	expressed sequence tags (SESTRs), useful for treating various disorders -
PT	such as autoimmune, infectious, and central nervous system disorders -
XX	
XX	
XX	Claim 1: Page 738; 803pp; English.
XX	
CC	AAA4426 to AAA45925 represent specifically claimed secreted expressed
CC	sequence tags (SESTRs), isolated from human, mouse, chicken and rat
CC	tissue sources. The SESTRs can have a range of activities depending on
CC	the tissues they are isolated from. The activities include:
CC	chemotactic; proliferative; immunomodulatory; haematopoietic;
CC	chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;

CC cytosolic; antibacterial; antifungal; antiviral; antidiabetic;
CC antistimetic; vulnereary; antitumor; osteopathic; neuroprotective;
CC nontoxic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC anticonvulsant; and antidepressant. The SESTs can be used for gene
CC therapy and in vaccines. The SESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumors, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 102 BP; 34 A; 17 C; 27 G; 24 T; 0 other;

Query Match 5.38; Score 68; DB 21; Length 102;
Best Local Similarity 100.0%; Pred. No. 4.5e-24;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 407 AATGATTAACATCTGCTGTCGTGACATGACCAATGATGTCGTAATTGGA 466
DB 27 aatgataaaacatactgctgtgtgtgacaaagcaatgtgtgtgaattgga 86
OY 467 AGCAAGCT 474
DB 87 agcaagct 94

RESULT 15
AA122766/c
ID AA122766 standard; DNA; 186 BP.
XX
AC AA122766;
DT 12-OCT-2001 (first entry)
XX
DE Probe #12699 for gene expression analysis in human cervical cell sample.
XX
KM Probe: human; microarray; gene expression; cervical epithelial cell;
KM cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN MO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001MO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 12699; 487pp; English.

XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human Hela cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 186 BP; 47 A; 44 C; 33 G; 62 T; 0 other;

Query Match 5.28; Score 66; DB 22; Length 186;
Best Local Similarity 100.0%; Pred. No. 4.5e-23;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 326 TATGTCGAATTGTTCCGATATAGATATATGAGACTATGTCGATCCATCCCTCAG 385
DB 147 TATGTCGAATTGTTCCGATATAGATATATGAGACTATGTCGATCCATCCCTCAG 88
OY 386 ACTTGG 391
DB 87 ACTTGG 82

RESULT 16
AA148068/c
ID AA148068 standard; DNA; 186 BP.
XX
AC AA148068;
DT 17-OCT-2001 (first entry)
XX
DE Probe #16754 used to measure gene expression in human placenta sample.
XX
KM Probe: microarray; human; placenta; antenatal diagnosis;
KM genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN MO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001MO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 16754; 654pp; English.

CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.

XX Sequence 186 BP; 47 A; 44 C; 33 G; 62 T; 0 other;

Query Match 5.2%; Score 66; DB 22; Length 186;
Best Local Similarity 100.0%; Pred. No. 4.5e-23;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

326 TATGTTGCGAATTTGTTCCCGTATAAAGATATATCTGGAAGTATGTCGCCATCCCTCAG 385
147 TATGTTGCGAATTTGTTCCCGTATAAAGATATATCTGGAAGTATGTCGCCATCCCTCAG 88

OY 386 ACTTGG 391
Db 87 ACTTGG 82

RESULT 17
AA108439/C
ID AA108439 standard; DNA; 186 BP.

XX AA108439;
AC
XX 09-OCT-2001 (first entry)

DE Probe #8430 used to measure gene expression in human breast sample.

XX Probe: human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

PN W0200157270-A2.

XX 09-AUG-2001.

PF 29-JAN-2001; 2001WO-US00661.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-476286/51.

XX Novel single exon nucleic acid probe used to measuring gene expression
in a human breast

PS Claim 25; SEQ ID No 8430; 322pp; English.

XX The present invention relates to novel single exon nucleic acid probes.

XX The present sequence is one such probe. The probes are useful for

XX measuring human gene expression in a human breast sample, where the probe

XX hybridises at high stringency to a nucleic acid expressed in the human

XX breast. The probes are useful for predicting, diagnosing, grading,

XX staging, monitoring and prognosing diseases of the human breast,

XX particularly those diseases with polygenic aetiology. The diseases

XX include: breast cancer, disorders of development, inflammatory diseases

XX of the breast, fibrocystic changes, proliferative breast disease, and

XX non-carcinoma tumours.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 186 BP; 47 A; 44 C; 33 G; 62 T; 0 other;

Query Match 5.2%; Score 66; DB 22; Length 186;
Best Local Similarity 100.0%; Pred. No. 4.5e-23;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 326 TATGTTGCGAATTTGTTCCCGTATAAAGATATATCTGGAAGTATGTCGCCATCCCTCAG 385
147 TATGTTGCGAATTTGTTCCCGTATAAAGATATATCTGGAAGTATGTCGCCATCCCTCAG 88

OY 386 ACTTGG 391
Db 87 ACTTGG 82

RESULT 18
AA113542/C
ID AA113542 standard; DNA; 411 BP.

XX AA113542;
AC
XX 12-OCT-2001 (first entry)

DE Probe #3475 for gene expression analysis in human cervical cell sample.

XX Probe: human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.

XX Homo sapiens.

PN W0200157278-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00670.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-489901/53.

XX Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human cervical epithelial cells

PS Claim 25; SEQ ID No 3475; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes

XX (SENP). The present sequence is one such probe. The SENPs are derived

XX from human HeLa cells. The SENPs can be used to produce a single exon

XX microarray, which can be used for measuring human gene expression in a

XX sample derived from human cervical epithelial cells. By measuring gene

XX expression, the probes are therefore useful in grading and/or staging

XX of diseases of the cervix, notably cervical cancer.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 411 BP; 106 A; 101 C; 67 G; 137 T; 0 other;

OY 326 TATGTTGCGAATTTGTTCCCGTATAAAGATATATCTGGAAGTATGTCGCCATCCCTCAG 385

```

DB 358 TATGTTGCGAATTTGTCCTCCGTAAAGATATATCTGGAACATGATGTCATCCCTCAG 299
QY 386 ACTTGG 391
    |||||
DB 298 ACTTGG 293

RESULT 19
ID AA134904/c
AC AA134904 standard; DNA: 411 BP.
XX
XX
XX 17-OCT-2001 (first entry)
XX
XX
XX Probe #3590 used to measure gene expression in human placenta sample.
DE
XX
XX Probe: microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-48897/53.
XX
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
XX
XX Claim 25; SEQ ID No 3590; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
XX
XX Sequence 411 BP; 106 A; 101 C; 67 G; 137 T; 0 other;
SQ

Query Match          5.2%; Score 66; DB 22; Length 411;
Best Local Similarity 100.0%; Pred. No. 4.6e-23;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 TATGTTGCGAATTTGTCCTCCGTAAAGATATATCTGGAACATGATGTCATCCCTCAG 385
    |||||
DB 358 TATGTTGCGAATTTGTCCTCCGTAAAGATATATCTGGAACATGATGTCATCCCTCAG 299
QY 386 ACTTGG 391
    |||||
DB 298 ACTTGG 293

RESULT 20
ID AA103431/c
AC AA103431 standard; DNA: 411 BP.

```

```

XX
XX AA103431;
AC
XX 09-OCT-2001 (first entry)
XX
XX
XX Probe #3422 used to measure gene expression in human breast sample.
DE
XX
XX Probe: human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
OS
XX
XX WO200157270-A2.
XX
XX 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US00661.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-476286/51.
XX
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
XX
XX Claim 25; SEQ ID No 3422; 322bp; English.
XX
XX
XX The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast.
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
XX Sequence 411 BP; 106 A; 101 C; 67 G; 137 T; 0 other;
SQ

Query Match          5.2%; Score 66; DB 22; Length 411;
Best Local Similarity 100.0%; Pred. No. 4.6e-23;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 TATGTTGCGAATTTGTCCTCCGTAAAGATATATCTGGAACATGATGTCATCCCTCAG 385
    |||||
DB 358 TATGTTGCGAATTTGTCCTCCGTAAAGATATATCTGGAACATGATGTCATCCCTCAG 299
QY 386 ACTTGG 391
    |||||
DB 298 ACTTGG 293

RESULT 21
ID AAH35811
AC AAH35811 standard; cDNA: 342 BP.

```


RESULT 23
AAZ52867 standard; cDNA: 1202 BP.
XX
XX
AC AAZ52867;
XX
XX
DT 14-MAR-2000 (first entry)
XX
DE Human prostate tumor cDNA library derived EST fragment #10.
XX
KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
KW treatment; ds.
XX
XX Homo sapiens.
OS
PN DE19820190-A1.
XX
PD 04-NOV-1999.
XX
PF 28-APR-1998; 98DE-1020190.
XX
PR 28-APR-1998; 98DE-1020190.
XX
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
DR WPI: 1999-621386/54.
DR P-PSDB: AAY73841, AAY73842, AAY73843.
XX
XX New human nucleic acid sequences from pancreatic tumors, and related
XX proteins.
PS Claim 2; Page 189-190; 502pp; German.
XX
XX This invention describes novel polypeptides and their encoding nucleic
XX acids derived from human pancreatic tumor tissue which have cytostatic
XX activity. The sequences are also useful in producing pharmaceutical
XX compositions for treatment of pancreatic tumors. AAZ52858-253014
XX represent expressed sequence tag (EST) fragments derived from a human
XX pancreatic tumor cDNA library and which encode the proteins represented
XX in AAY73814-Y74252.
CC
CC
SQ Sequence 1202 BP; 381 A; 227 C; 267 G; 327 T; 0 other;
XX
XX
Query Match 1.5%; Score 19; DB 20; Length 1202;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 378 TCCCTCAGACTTGGGAGA 396
DB 454 tccctcagacttgggaaga 472
XX
XX
RESULT 24
AAC98069
ID AAC98069 standard; cDNA: 1215 BP.
XX
XX AAC98069;
AC
XX
DT 09-MAR-2001 (first entry)
XX
XX Human colon cancer antigen nucleotide sequence SEQ ID NO:79.
DE
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder; ss.
XX
XX

OS Homo sapiens.
XX
XX WO200055351-A1.
PN
XX
XX 21-SEP-2000.
PD
XX
XX 08-MAR-2000; 2000WO-US05883.
PF
XX
XX 12-MAR-1999; 99US-0124270.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ruben SM;
PI
XX
XX WPI: 2000-587534/55.
DR
XX
XX P-PSDB: AAB53312.
DR
XX
XX Colon cancer associated gene sequences, referred to as colon cancer
XX PT antigens, useful for the treatment, prevention, and diagnosis of colon
XX PT disorders such as colon cancer.
PS
XX
XX Claim 1; Page 524; 2104pp; English.
XX
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
XX CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
XX CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
XX CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
XX CC vulnary, nephrotropic, antiinfective and antibacterial activities, and
XX CC can be used in gene therapy. The colon cancer antigen polynucleotides,
XX CC proteins and antibodies to the proteins are useful for the prevention,
XX CC treatment and diagnosis of colon disorders, such as colon cancer. The
XX CC polynucleotides may be used in diagnostics and research, such as for
XX CC chromosome identification, and as hybridisation probes. The proteins
XX CC may also be used to prevent diseases such as neural disorders, immune
XX CC system disorders, muscular disorders, reproductive disorders,
XX CC gastrointestinal disorders, wounds, renal disorders, infectious
XX CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
XX CC AAB54007 represent sequences used in the exemplification of the present
XX CC invention.
XX
XX
SQ Sequence 1215 BP; 381 A; 232 C; 273 G; 329 T; 0 other;
XX
XX
Query Match 1.5%; Score 19; DB 21; Length 1215;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 378 TCCCTCAGACTTGGGAGA 396
DB 454 tccctcagacttgggaaga 472
XX
XX
RESULT 25
AAI59678
ID AAI59678 standard; cDNA: 1536 BP.
XX
XX AAI59678;
AC
XX
DT 22-OCT-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 3667.
DE
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
PN
XX

PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
XX P-PSDB: AAM40522.
PT Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 3667; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1536 BP; 443 A; 365 C; 399 G; 329 T; 0 other;

Query Match 1.5%; Score 19; DB 22; Length 1536;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 GTGCAGCGCGCGCGCAG 76
|||||
Db 346 gtgcagcgcgcgcgcgag 364

RESULT 26
ID AA159679 standard; cDNA: 1536 BP.
XX
AC AA159679;
XX
DT 22-OCT-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 3668.
XX
XX Human, nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX

OS Homo sapiens.
XX
XX W0200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
XX P-PSDB: AAM40523.
PT Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 3668; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1536 BP; 443 A; 365 C; 399 G; 329 T; 0 other;

Query Match 1.5%; Score 19; DB 22; Length 1536;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 GTGCAGCGCGCGCGCAG 76
|||||
Db 346 gtgcagcgcgcgcgcgag 364

RESULT 27
ID AAV81267 standard; cDNA: 2264 BP.
XX
AC AAV81267;
XX
DT 11-MAR-1999 (first entry)
XX
XX Bcl-2 interaction protein encoding cDNA.
XX Bcl-2 interaction protein; Bcl-2; molecular mechanism; apoptosis;
KW diagnosis; prevention; treatment; ss.
XX

OS Homo sapiens.
 XX Key Location/Qualifiers
 FH CDS 213..1742
 FT /*tag= a
 FT /product= "Bcl-2 interaction protein"
 XX
 XX JP10327872-A.
 XX
 PD 15-DEC-1998.
 XX
 XX 30-MAY-1997; 97JP-0157841.
 XX
 XX 30-MAY-1997; 97JP-0157841.
 XX
 XX 30-MAY-1997; 97JP-0157841.
 XX
 XX (SAKA) OTSUKA PHARM CO LTD.
 XX
 XX WPI: 1999-099033/09.
 XX
 XX P-PSDB: AAM95196.
 XX
 XX New Bcl-2 interaction protein gene (Bis) - useful for elucidation of
 PT the molecular mechanism of apoptosis, and in diagnosis, prevention
 PT and treatment of diseases
 PS Claim 2; Page 12-15; 19pp; Japanese.
 XX
 CC This cDNA encodes a Bcl-2 interaction protein which has combining
 CC activity to Bcl-1. The Bcl-2 interaction protein and its gene (Bis) are
 CC useful for the elucidation of the molecular mechanism of apoptosis and
 CC various diseases related to it, and their diagnosis, prevention and
 CC treatment.
 XX
 SO Sequence 2264 BP; 565 A; 687 C; 574 G; 437 T; 1 other;

Query Match 1.5%; Score 19; DB 20; Length 2264;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 58 GTGGCAGCGCGCGCGCAG 76
 ||||||||||||||||||
 DB 474 gtgcaagcgcgcgcgcgag 492

RESULT 28
 AA251797
 ID AA251797 standard; cDNA: 2528 BP.
 XX
 XX AC AA251797;
 XX
 DT 04-JUL-2000 (first entry)
 XX
 DE Full length cDNA sequence for human BAG-3 protein.
 XX
 KW Human BAG-3; Bcl-2 associated athanogene-3; apoptosis; cell migration;
 KW tumour cell metastasis; Hsc70/Hsp70-regulating protein; metastasis;
 KW tumour cell proliferation; steroid hormone receptor function; ss.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH CDS 1..2034
 FT /*tag= a
 FT /product= "Human BAG-3 protein"
 XX
 XX WO200014106-A1.
 XX
 PD 16-MAR-2000.
 XX
 XX 09-SEP-1999; 99WO-US21053.
 XX
 XX 09-SEP-1998; 98US-0150489.
 XX

PA (BURN-) BURNHAM INST.
 XX
 XX Reed JC, Takayama S;
 PI
 XX
 DR WPI: 2000-256937/22.
 XX
 DR P-PSDB: AAY70508.
 XX
 XX BAG-1 related proteins from humans, Caenorhabditis elegans and
 PT Schizosaccharomyces pombe useful for modulating tumor cell
 PT proliferation, cell migration and metastasis and steroid hormone
 PT receptor function
 PS Claim 7; Fig 3; 132pp; English.
 XX
 CC The present cDNA sequence encodes human BAG-3 (Bcl-2 associated
 CC athanogene-3) protein. BAG is a Hsc70/Hsp70-regulating protein
 CC (Hsc70/Hsp70 is a molecular chaperone that participates in controlling
 CC protein bioactivity, degradation, complex assembly/disassembly and
 CC translocation across membranes). It competes with Hsp for binding to the
 CC Hsc70/Hsp70 ATPase binding domain and promotes substrate release. Gene
 CC transfection studies indicate that BAG proteins influence a wide variety
 CC of cellular phenotypes through their interactions with Hsc70/Hsp70,
 CC including increasing resistance to apoptosis, promoting cell
 CC proliferation, enhancing tumour cell migration and metastasis and
 CC altering transcriptional activity of steroid hormones. The ability of
 CC BAG-2 and BAG-3 proteins to interact in cells with Hsc70 was tested by
 CC expressing these proteins with N-terminal Flag group tags in 293T human
 CC epithelial cells using immunoprecipitation assays. BAG also stimulates
 CC Hsc70-mediated adenosine triphosphate (ATP) hydrolysis by accelerating
 CC ADP/ATP exchange.
 XX
 SO Sequence 2528 BP; 586 A; 805 C; 656 G; 478 T; 3 other;

Query Match 1.5%; Score 19; DB 21; Length 2528;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 58 GTGGCAGCGCGCGCGCAG 76
 ||||||||||||||||||
 DB 766 gtgcaagcgcgcgcgcgag 784

RESULT 29
 AA251804
 ID AA251804 standard; cDNA: 2534 BP.
 XX
 XX AC AA251804;
 XX
 DT 04-JUL-2000 (first entry)
 XX
 DE Full length expanded cDNA sequence for human BAG-3 protein.
 XX
 KW Human BAG-3; Bcl-2 associated athanogene-3; apoptosis; cell migration;
 KW tumour cell metastasis; Hsc70/Hsp70-regulating protein; metastasis;
 KW tumour cell proliferation; steroid hormone receptor function; ss.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH CDS 307..2034
 FT /*tag= a
 FT /product= "Human BAG-3 protein"
 XX
 XX WO200014106-A1.
 XX
 PD 16-MAR-2000.
 XX
 XX 09-SEP-1999; 99WO-US21053.
 XX
 XX 09-SEP-1998; 98US-0150489.
 XX
 XX (BURN-) BURNHAM INST.
 PA

CC transformation of a eukaryotic cell comprises: (a) transforming or
CC transfecting the cell with a vector carrying a nucleic acid encoding
CC a mutated (Na,K)-ATPase gene of the invention, plus a gene of
CC interest; (b) incubating the cells with a cardioglycoside such
CC as digitoxin, digitogenin, digoxin, bufalin or (especially) ouabain;
CC (c) contacting the incubated cells with a fibrinogen-coated solid
CC matrix; and (d) detecting any cells bound to the matrix. The new
CC selectable marker is particularly useful where antibiotic resistance
CC selection is undesirable, e.g. in gene therapy.

XX
SQ Sequence 3636 BP; 850 A; 967 C; 995 G; 824 T; 0 other;

Query Match 1.5%; Score 19; DB 20; Length 3636;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GTGGCAGCGCGCGCGCAG 76
|||||
DB 28 gtgcagcgcgcgcgcgcag 46

RESULT 32

AAV84296
ID AAV84296 standard; DNA; 3636 BP.

XX
AC AAV84296;

XX
DT 12-APR-1999 (first entry)

XX
DE Rat (Na,K)-ATPase alpha 1 subunit (L879C, L881C, mutant) DNA.

XX
KM Selectable marker; (Na,K)-ATPase; rat; gene therapy; ss.

XX
OS Rattus rattus.

XX
OS Synthetic.

XX
PN WO9855603-A1.

XX
PD 10-DEC-1998.

XX
PF 04-JUN-1998; 98WO-SE01062.

XX
PR 04-JUN-1997; 97SE-0002120.

XX
PA (KARO-) KAROLINSKA INNOVATIONS AB.

XX
PI Belusa R;

XX
DR WPI: 1999-059908/05.

XX
P-PSDB; AAM88289.

XX
PT New mutated (Na,K)-ATPase - useful as a selection marker

XX
PS Disclosure; Page 54-58; 72pp; English.

CC This nucleotide sequence encodes a new mutated (Na,K)-ATPase alpha
CC 1 subunit (see AAM88288), in which the Leu-879 and Leu-881 residues of
CC the native rat protein (see AAM88290) have been mutated to cysteine.
CC This mutated protein can be used as a novel selectable marker. The
CC mutation, obtained by mutagenesis of the (Na,K)-ATPase gene (see
CC AAV84297), causes a loss in sensitivity of the encoded enzyme to
CC ouabain. A claimed method for detecting transfection or
CC transformation of a eukaryotic cell comprises: (a) transforming or
CC transfecting the cell with a vector carrying a nucleic acid encoding
CC a mutated (Na,K)-ATPase gene of the invention, plus a gene of
CC interest; (b) incubating the cells with a cardioglycoside such
CC as digitoxin, digitogenin, digoxin, bufalin or (especially) ouabain;
CC (c) contacting the incubated cells with a fibrinogen-coated solid
CC matrix; and (d) detecting any cells bound to the matrix. The new
CC selectable marker is particularly useful where antibiotic resistance
CC selection is undesirable, e.g. in gene therapy.

SQ Sequence 3636 BP; 850 A; 966 C; 995 G; 825 T; 0 other;

Query Match 1.5%; Score 19; DB 20; Length 3636;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GTGGCAGCGCGCGCGCAG 76
|||||
DB 28 gtgcagcgcgcgcgcgcag 46

RESULT 33

AAV84297
ID AAV84297 standard; DNA; 3636 BP.

XX
AC AAV84297;

XX
DT 12-APR-1999 (first entry)

XX
DE Rat (Na,K)-ATPase alpha 1 subunit gene coding region.

XX
KM Selectable marker; (Na,K)-ATPase; rat; gene therapy; ss.

XX
OS Rattus rattus.

XX
PN WO9855603-A1.

XX
PD 10-DEC-1998.

XX
PF 04-JUN-1998; 98WO-SE01062.

XX
PR 04-JUN-1997; 97SE-0002120.

XX
PA (KARO-) KAROLINSKA INNOVATIONS AB.

XX
PI Belusa R;

XX
DR WPI: 1999-059908/05.

XX
P-PSDB; AAM88290.

XX
PT New mutated (Na,K)-ATPase - useful as a selection marker

XX
PS Disclosure; Page 59-63; 72pp; English.

CC This nucleotide sequence encodes the rat (Na,K)-ATPase alpha 1
CC subunit (see AAM88290). (Na,K)-ATPase is an ion transporter
CC essential to the cell. The invention provides new, mutated
CC (Na,K)-ATPases (see AAM88287-89) that can be used as novel
CC selectable markers. Appropriate mutation of the (Na,K)-ATPase
CC gene (see AAV84294-96) causes a loss in sensitivity of the encoded
CC enzyme to ouabain. A claimed method for detecting transfection or
CC transformation of a eukaryotic cell comprises: (a) transforming or
CC transfecting the cell with a vector carrying a nucleic acid
CC encoding a mutated (Na,K)-ATPase gene of the invention, plus a gene
CC of interest; (b) incubating the cells with a cardioglycoside such
CC as digitoxin, digitogenin, digoxin, bufalin or (especially) ouabain;
CC (c) contacting the incubated cells with a fibrinogen-coated solid
CC matrix; and (d) detecting any cells bound to the matrix. The new
CC selectable markers are particularly useful where antibiotic
CC resistance selection is undesirable, e.g. in gene therapy.

XX
SQ Sequence 3636 BP; 850 A; 968 C; 995 G; 823 T; 0 other;

Query Match 1.5%; Score 19; DB 20; Length 3636;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GTGGCAGCGCGCGCGCAG 76
|||||
DB 28 gtgcagcgcgcgcgcgcag 46

RESULT 34

AAAF44683 standard; cDNA: 7328 BP.

AAAF44683;

27-MAR-2001 (first entry)

Novel protein kinase cDNA, SEQ ID NO: 63.

Human; mouse; protein kinase; antiarthritis; antisclerotic; osteopathic; immunosuppressive; cardiant; renal; antiinflammatory; antistimatic; dermatological; antidiabetic; antineoplastic; gene therapy; vaccine; immune disorder; cardiovascular disease; neurodegenerative disease; cancer; autoimmune disorder; stroke; inflammatory bowel disease; inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.

Homo sapiens.

MO200073469-A2.

07-DEC-2000.

26-MAY-2000. 2000MO-US14842.

28-MAY-1999. 99US-0136503.

(SUGC-) SUGEN INC.

Plowman GD, Martinez R, Whyte D, Sudersanam S;

WPI: 2001-032161/04.

P-PSDB: AAB65656.

Nucleic acids encoding kinase polypeptides, useful for diagnosing and treating immune-related diseases and disorders, cardiovascular disease, neurodegenerative diseases and/or cancers -

Example 1; Fig 2; 310pp; English.

The present sequence encodes a novel protein kinase. The nucleic acids and the protein kinases they encode may be used in the treatment and diagnosis of diseases associated with inappropriate kinase expression such as immune-related diseases and disorders, cardiovascular disease, neurodegenerative diseases and/or cancers. The nucleic acids and complementary sequences may also be used as DNA probes in diagnostic assays. The kinase polypeptides may be used as antigens in the production of antibodies of kinase expression and activity. Anti-kinase antibodies and kinase antagonists may also be used to down regulate kinase expression and activity. Diseases related to kinase expression and activity include rheumatoid arthritis, atherosclerosis, autoimmune disorders, complications of organ transplantation, myocardial infarction, immune disorders, cardiomyopathies, strokes, renal failure, oxidative-stress related disorders, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and reproductive disorders.

Sequence 7328 BP: 2036 A; 1850 C; 1655 G; 1785 T; 2 other;

Query Match

Best Local Similarity 100.0%; Score 19; DB 22; Length 7328; Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GTGGCAGCGCGCGCGCAG 76

DB 589 GTGGCAGCGCGCGCGCAG 607

RESULT 35

AAZ38983

AAZ38983 standard; DNA: 30 BP.

AAZ38983;

28-FEB-2000 (first entry)

Expression vector construction primer SEQ ID NO:18.

1,3-propanediol; vitamin B12 transport; glycerol dehydratase; BluC; 1,3-propanediol oxidoreductase; vitamin B12 receptor precursor; BluC; vitamin B12 transport system permease protein; Blud; polyester fibre; vitamin B12 transport ATP-binding protein; polyurethane; primer; ss.

Synthetic.

W09958686-A2.

18-NOV-1999.

12-MAY-1999. 99MO-US10356.

12-MAY-1998. 98US-0085190.

(DUPO) DU PONT DE NEMOURS & CO E I.

(GENV) GENENCOR INT INC.

Whited GM, Bulthuis B, Trimbur DE, Gatenby AA;

WPI: 2000-053104/04.

production of a monomer which is used in the manufacture of polyester fiber, polyurethane and cyclic compounds -

Example 1; Page 56; 61pp; English.

The present invention describes the monomer, 1,3-propanediol (I), which is biologically produced by using a transformed host cell (C) comprising at least one copy of gene encoding Blud, Bluc and Blud. (I) is produced by the steps comprising: (1) contacting the transformed (C) with at least one fermentable common source and an effective amount of vitamin B12. The one transformed (C) comprises at least one copy of genes encoding a protein having dehydratase activity (a), a protein having an oxidoreductase activity (b), a vitamin B12 receptor precursor protein (c), a vitamin B12 transport system permease protein (d), and a vitamin B12 transport ATP or GTP-binding protein (e); and (2) recovering (I) produced by (1). The method is used to produce (I) which is employed in the manufacture of polyester fibre, polyurethane and cyclic compounds. The present invention produces 1,3-propanediol rapidly and it is inexpensive. The present sequence represents a primer used in the construction of expression vectors in the exemplification of the present invention.

Sequence 30 BP: 8 A; 4 C; 6 G; 12 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 18; DB 21; Length 30; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 872 ATGTGAGCTTACCA 889

DB 4 attgtgagcttaccac 21

RESULT 36

AAV90298 standard; cDNA: 346 BP.

AAV90298;

15-FEB-1999 (first entry)

EST clone D1482.

KM Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KM receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KM gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO9845436-A2.
XX
PD 15-OCT-1998.
XX
PE 10-APR-1998; 98WO-US06955.
XX
PR 10-APR-1997; 97US-0838821.
XX
PA (GEMV) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1999-070077/06.
XX
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries.
XX
PS Claim 1; Page 496; 618pp; English.
XX
CC The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.
XX
SQ Sequence 346 BP; 128 A; 62 C; 101 G; 55 T; 0 other;

Query Match 1.4%; Score 18; DB 20; Length 346;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 680 GGAACACGAAATGAG 697
|||||
DB 266 ggaacacgagaatgag 283

RESULT 37
AAC01530
ID AAC01530 standard; cDNA; 402 BP.
XX
AC AAC01530;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 1528.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.

XX
XX 26-FEB-1999; 99US-0122487.
PR
PA (GEMV) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PDB; AAC01524.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 1528; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 402 BP; 144 A; 68 C; 113 G; 70 T; 7 other;

Query Match 1.4%; Score 18; DB 21; Length 402;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 680 GGAACACGAAATGAG 697
|||||
DB 244 ggaacacgagaatgag 261

RESULT 38
AAF11454/C
ID AAF11454 standard; cDNA; 543 BP.
XX
AC AAF11454;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus niger EST SEQ ID NO:3977.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus niger.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;

XX DR WPI: 2000-594572/56.
 XX PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX PS Claim 87; Page 1764; 3161pp; English.

CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random DNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AA07478 to AA11247 represents ESTs from
 CC *Aspergillus venenatum*; AA11248 to AA11853 represents ESTs from *Aspergillus*
 CC *fligeri*; AA11854 to AA14878 represents ESTs from *Aspergillus oryzae*; and
 CC AA14879 to AA15337 represents ESTs from *Trichoderma reesei*, which are
 CC all specifically claimed in the present invention.

XX SO Sequence 543 BP; 128 A; 128 C; 121 G; 166 T; 0 other;

Query Match 1.4%; Score 18; DB 21; Length 543;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 GGCACGCGCGCGCGCAG 77
 DB 163 GGCACGCGCGCGCGCAG 146

RESULT 39
 AAH13135/C
 ID AAH13135 standard; cDNA; 582 BP.
 AC AAH13135;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (3'-primer); SEQ ID NO:9970.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.

XX XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI: 2001-318749/34.
 DR
 XX
 PT primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 XX Claim 3; SEQ ID 9970; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets are useful for synthesizing polynucleotides,
 CC in gene therapy. The primers are also useful for the
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX SO Sequence 582 BP; 142 A; 120 C; 83 G; 229 T; 8 other;

Query Match 1.4%; Score 18; DB 22; Length 582;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 870 CCATTGTGATGCTTTAC 887
 DB 106 CCATTGTGATGCTTTAC 89

RESULT 40
 AAH67523/C
 ID AAH67523 standard; DNA; 588 BP.
 AC AAH67523;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum coding sequence fragment SEQ ID NO: 2558.
 XX
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX

PA (UYHE-) UNIV HELSINKI LICENSING LTD OY.
XX Alltalo K, Eriksson U, Olofsson B, Pajusola K;
XX WPI: 1996-412582/41.
DR P-PSDB: AAM04831.
XX Vascular endothelial growth factor VEGF-B proteins - useful to
PT accelerate angiogenesis in wound healing, also related nucleic acid
PT and antibodies for cancer diagnosis
XX
XX Claim 1: Page 61-62; 107pp; English.
XX AA17909-17915 represent the coding sequences for the vascular
CC endothelial growth factor (VEGF) proteins of the invention, which promote
CC endothelial or mesodermal cell proliferation. VEGF is also a
CC glycosylated cationic dimer, and is sometimes referred to as vascular
CC permeability factor (VPF). VEGF has diverse effects, depending on the
CC specific biological context in which it is found. VEGF is a potent
CC endothelial cell mitogen, and directly contributes to induction of
CC angiogenesis in vivo by promoting endothelial cell growth during normal
CC angiogenic development, wound healing, and tissue
CC embryonic development. The VEGF proteins of the invention share
CC regenerative/reorganisational. The VEGF proteins can therefore be
CC the angiogenic and other properties of VEGF. The proteins can therefore be
CC expressed in tissues differently to VEGF. The proteins can therefore be
CC used to accelerate angiogenesis in wound healing. Antibodies against the
CC proteins can be used for inhibiting angiogenesis. The antibodies can
CC also be used diagnostically to quantitatively detect the invention of
CC complementary to the coding sequences for the proteins of the invention of
CC can also be used to detect VEGF-B coding sequences. Quantification of
CC VEGF-B in cancer biopsy specimens may be useful as an indicator of
CC metastatic risk. VEGF-B expression in a cell can be retarded using
CC antisense sequences direct against the VEGF coding sequences, this is
CC especially useful in retarding VEGF expression in tumour cells.
SQ Sequence 624 BP; 106 A; 235 C; 179 G; 104 T; 0 other;

Query Match 1.4%; Score 18; DB 17; Length 624;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 GGCAGCGCGCGCGCAGG 77
DB 585 GGCAGCGCGCGCGCAGG 568

RESULT 43
AAV63569/c
ID AAV63569 standard; cDNA: 624 BP.
XX
XX AAV63569;
XX 29-JAN-1999 (first entry)
XX
XX cDNA clone encoding vascular endothelial growth factor (VEGF)-B186.
XX
XX Vascular endothelial growth factor; VEGF: proliferation; human;
XX endothelial cell; angiogenesis; tissue growth; organ repair; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..624
XX FT /*tag= a
XX FT /product= VEGF-B186
XX
XX US5840693-A.
XX
XX 24-NOV-1998.
XX
XX 01-MAR-1996; 96US-0609443.

PR 01-MAR-1996; 96US-0609443.
PR 01-MAR-1995; 95US-0397651.
PR 06-JUN-1995; 95US-0469427.
PR 06-DEC-1995; 95US-0569063.
XX
XX (LUDWIG) LUDWIG INST CANCER RES.
XX (UYHE-) UNIV HELSINKI LICENSING LTD OY.
XX Alltalo K, Eriksson U, Olofsson B, Pajusola K;
XX WPI: 1999-034079/03.
DR P-PSDB: AAM80495.
XX Vascular endothelial growth factor-B isoforms, and DNA encoding
PT them, useful for inducing angiogenesis and cellular proliferation,
PT and raising antibodies to inhibit activities in e.g. tumours
XX
XX Claim 22; Fig 14; 52pp; English.
XX The present sequence encodes human vascular endothelial growth factor
CC (VEGF)-B186. VEGF proteins are used for promoting proliferation of
CC endothelial cells and for stimulating angiogenesis (the proliferation
CC of new capillaries form pre-existing blood vessels). These activities
CC are useful for treating tissue growth and repair, including organ
CC repair. This is also useful in pregnancy, in follicle development,
CC as these processes must occur in development of the placenta. The
CC proteins can also be used to raise antibodies, either for use in
CC detection of the proteins or as inhibitors of their action. This is
CC especially useful as angiogenesis is required by tumours as they need
CC new blood supplies to grow and proliferate.
SQ Sequence 624 BP; 106 A; 235 C; 179 G; 104 T; 0 other;

Query Match 1.4%; Score 18; DB 20; Length 624;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 GGCAGCGCGCGCGCAGG 77
DB 585 GGCAGCGCGCGCGCAGG 568

RESULT 44
AAA91005/c
ID AAA91005 standard; DNA: 624 BP.
XX
XX AAA91005;
XX 05-APR-2001 (first entry)
XX
XX Human VEGF-3 coding sequence.
XX
XX Human; angiogenic protein; wound healing; vascular tissue repair;
XX peripheral arterial disease; critical limb ischaemia; coronary disease;
XX angiogenesis; tumour; inflammation; diabetic retinopathy; psoriasis;
XX rheumatoid arthritis; autoimmune disease; allergy; cancer; therapy;
XX infectious disease; neurodegeneration;
XX vascular endothelial growth factor-3; VEGF-3; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..624
XX FT /*tag= a
XX FT /product= VEGF-3
XX
XX WO200075163-A1.
XX
XX 14-DEC-2000.
XX
XX 01-JUN-2000; 2000WO-US14925.

PR 03-JUN-1999; 9905-0137796.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Hu J, Cao L;
XX WPI; 2001-071057/08.
DR P-PSDB; AAY97571.
XX
PT New nucleic acid encoding angiogenic proteins, useful e.g. for
PT promoting healing of wounds and treating peripheral arterial disease,
PT critical limb ischemia or coronary disease
PS
PS Claim 1; Page 224; 244pp; English.
XX
CC This sequence encodes vascular endothelial growth factor-3 (VEGF-3),
CC which is an angiogenic protein of the invention. The angiogenic proteins
CC and the DNA sequences encoding them, are used to prevent, treat or
CC ameliorate disease and to detect diseases, or susceptibility, by
CC detecting mutations or the presence or amount of angiogenic protein
CC expression. Particularly they are used to stimulate wound healing,
CC growth of damaged bone and tissue, and for repair of vascular tissue,
CC especially peripheral arterial disease, critical limb ischemia or
CC coronary disease. Antagonists of the sequences are used to inhibit
CC angiogenesis in tumours and to treat inflammation (where associated with
CC increased vascular permeability), diabetic retinopathy, rheumatoid
CC arthritis or psoriasis. Agonists are also useful for stimulating
CC (lymph)angiogenesis. The proteins are also used to identify specific
CC binding agents (potential therapeutic agents) and to raise antibodies.
CC The antibodies are useful as therapeutic (antagonists); for detection,
CC purification and targeting of proteins for in vivo or in vitro diagnosis
CC (including imaging) or for therapy (including when linked to e.g. a label
CC or cytotoxin); and for immunotyping of cells, e.g. for detecting minimal
CC residual disease or hematopoietic progenitor/stem cells. It is also
CC contemplated that the sequences might be useful for treating a very wide
CC range of other disorders, e.g. autoimmune diseases; allergy; cancer;
CC infectious diseases (viral, bacterial, fungal or parasitic);
CC neurodegeneration, also as chemotactic agents or for stimulating
CC regeneration of the nervous system etc.
XX
XX Sequence 624 BP; 106 A; 235 C; 179 G; 104 T; 0 other:
SO
Query Match 1.4%; Score 18; DB 22; Length 624;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 60 GGCAGCGCGCGCGCAGG 77
DB 585 GGCAGCGCGCGCGCAGG 568
RESULT 45
ID AAA68125 standard; DNA; 762 BP.
AC AAA68125;
XX
XX 24-OCT-2000 (first entry)
XX
DE Pinus radiata flavonol glucosyl transferase DNA sequence SEQ ID NO:218.
XX
XX Plant; Lignin; Lignin biosynthetic pathway; Eucalyptus grandis;
XX Pinus radiata; Monterey pine; ds.
XX
OS Pinus radiata.
XX
XX WO200022099-A1.
XX
XX 20-APR-2000.
XX
XX 06-OCT-1999; 99WO-N200168.
XX
XX

PR 09-OCT-1998; 9805-0169789.
PR 14-JUL-1999; 9905-0143811.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Bloksberg LN, Havukkala IU;
PI
DR WPI; 2000-317962/27.
XX
XX Novel polynucleotide encoding enzymes involved in lignin-biosynthetic
PT pathway useful for producing transgenic plants especially eucalyptus
PT and pine species having altered lignin content, composition and
PT structure
XX
PS Claim 1; Page 124; 213pp; English.
XX
XX The present invention describes isolated polynucleotides and proteins
CC encoding and representing the enzymes cinnamate 4-hydroxylase (C4H),
CC coumarate 3-hydroxylase (C3H), phenolase (PNL), O-methyl transferase
CC (OMT), cinnamyl alcohol dehydrogenase (CAD), cinnamoyl-CoA reductase
CC (CCR), phenylalanine ammonia-lyase (PAL), 4-coumarate-CoA ligase (4CL),
CC coniferol glucosyl transferase (CGT), coniferin beta-glucosidase (CBG),
CC laccase, peroxidase, ferulate-5-hydroxylase (F5H), alpha-amylase,
CC caffeic acid methyl transferase, caffeoyl CoA methyl transferase,
CC coumarate CoA ligase, cytochrome P450 1X1A, diphenol oxidase, flavanol
CC glucosyl transferase, flavenoid hydroxylase, and isoflavone reductase,
CC which are involved in the lignin biosynthetic pathway. The
CC polynucleotides can be used for modulating lignin content, lignin
CC composition and the structure of a plant, especially eucalyptus and pine
CC species, and for modifying the activity of an enzyme involved in lignin
CC biosynthetic pathway, and for producing a plant having altered lignin
CC content, composition and structure. They can be used for designing probes
CC and primers useful for detecting similar DNA and RNA sequences in any
CC organism and for PCR amplification. The lignin content can be efficiently
CC modified using the polynucleotides. AAA67908 to AAA68201 and AAB16341 to
CC AAB16449 represent polynucleotide and protein sequences used in the
CC exemplification of the present invention.
XX
XX Sequence 762 BP; 227 A; 131 C; 202 G; 202 T; 0 other:
SO
Query Match 1.4%; Score 18; DB 21; Length 762;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 764 CATGGAAGCATTCAGTC 781
DB 241 catggaagcattagtg 258

Search completed: March 29, 2002, 20:56:33
Job time: 2153 sec

Mon Apr 1 08:00:08 2002

us-09-415-540-2.oli.rng

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2002, 19:55:10 ; Search time 90.84 Seconds
(without alignments)
3178.770 Million cell updates/sec

Title: US-09-415-540-2

Perfect score: 1275

Sequence: 1 CAACAGGTTNGGCGCTCTCT.....GAATTAATTTTCTG 1275

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1275	100.0	1275	2	US-08-741-437-2
2	1275	100.0	1275	2	US-09-134-593-2
3	18	1.4	624	2	US-08-609-443B-14
4	18	1.4	624	2	US-08-569-063C-14
5	18	1.4	4825	6	5459251-1
6	18	1.4	5086	2	US-08-465-485A-19
7	18	1.4	5086	2	US-08-365-486A-14
8	18	1.4	5086	3	US-09-080-285-19
9	18	1.4	5086	4	US-08-880-342-14
10	18	1.4	5086	5	PCR-US93-05651-4
11	18	1.4	5086	5	PCR-US93-06251-2
12	18	1.4	5104	6	5306344-1
13	17	1.3	77	2	US-08-180-524-5
14	17	1.3	77	2	US-08-975-166-5
15	17	1.3	146	2	US-08-180-524-3
16	17	1.3	146	2	US-08-975-166-3
17	17	1.3	461	1	US-08-454-720A-37
18	17	1.3	617	4	US-09-328-111-577
19	17	1.3	718	3	US-09-154-083-31
20	17	1.3	810	3	US-09-097-072-1
21	17	1.3	900	3	US-08-886-638-1
22	17	1.3	900	3	US-09-097-072-3
23	17	1.3	900	4	US-09-212-988-1
24	17	1.3	1068	5	PCR-US91-00889-13
25	17	1.3	1215	5	PCR-US91-00889-5
26	17	1.3	1488	1	US-07-914-281-9
27	17	1.3	1488	1	US-08-393-246-9

28	17	1.3	1488	1	US-08-525-058A-9	Sequence 9, Appl
29	17	1.3	1488	2	US-08-696-731-9	Sequence 9, Appl
30	17	1.3	1488	4	US-09-042-531-9	Sequence 9, Appl
31	17	1.3	1846	2	US-08-365-486A-16	Sequence 16, Appl
32	17	1.3	1846	4	US-08-880-342-16	Sequence 16, Appl
33	17	1.3	2134	5	PCR-US96-06427-3	Sequence 3, Appl
34	17	1.3	2134	5	PCR-US96-06427-3	Sequence 3, Appl
35	17	1.3	2175	4	US-08-482-073-10	Sequence 10, Appl
36	17	1.3	2861	4	US-08-482-073-10	Sequence 10, Appl
37	17	1.3	3319	1	US-08-006-676B-2	Sequence 1, Appl
38	17	1.3	3319	1	US-08-282-845-1	Sequence 1, Appl
39	17	1.3	3319	2	US-08-428-414A-4	Sequence 4, Appl
40	17	1.3	3319	5	PCR-US94-00324-2	Sequence 4, Appl
41	17	1.3	3467	3	US-08-972-719-1	Sequence 1, Appl
42	17	1.3	3467	4	US-08-753-038-1	Sequence 1, Appl
43	17	1.3	3647	1	US-07-914-281-7	Sequence 7, Appl
44	17	1.3	3647	1	US-08-393-246-7	Sequence 7, Appl
45	17	1.3	3647	1	US-08-525-058A-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-08-741-437-2
Sequence 2, Application US/08741437
Patent No. 5843665
GENERAL INFORMATION:
APPLICANT: Hawkins, Phillip R.
TITLE OF INVENTION: A NOVEL HUMAN PYROPHOSPHATASE
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSER: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/741,437
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0148 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 855-0555
TELEFAX: (415) 845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1275 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY:
CLONE: Consensus
US-08-741-437-2
Query Match 100.0%; Score 1275; DB 2; Length 1275;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 CAAGAGTTTNGGGGCTCTCTCTTGTGAGTGGGCGCGGCTGTGCTGTG 60
DB 1 CAAGAGTTTNGGGGCTCTCTCTTGTGAGTGGGCGCGGCTGTGCTGTG 60
OY 61 GCAGGGGGGGGGGAGGAGTCCGGCATTATGAGCGGCTTACAGCAGAGGAGCGCGG 120
DB 61 GCAGGGGGGGGGGAGGAGTCCGGCATTATGAGCGGCTTACAGCAGAGGAGCGCGG 120
OY 121 GCGCTTCCCTGGAGTACCGAGTCTTCTCAAAATGAGAAAGACATATATCTCC 180
DB 121 GCGCTTCCCTGGAGTACCGAGTCTTCTCAAAATGAGAAAGACATATATCTCC 180
OY 181 ATTTCATATATTCATTTATGAGATAGAGATGTGTTACATGATAGTGAATACC 240
DB 181 ATTTCATATATTCATTTATGAGATAGAGATGTGTTACATGATAGTGAATACC 240
OY 241 ACGTGTCTATATGCAAAATGAGATGTCTACAAAGGACCTTTAAACCTATTAAACA 300
DB 241 ACGTGTCTATATGCAAAATGAGATGTCTACAAAGGACCTTTAAACCTATTAAACA 300
OY 301 AGATGTGAAAAAGGAAAACTTGTCTATGTGCGAATTTGTTCCGCTATTAAGATATAT 360
DB 301 AGATGTGAAAAAGGAAAACTTGTCTATGTGCGAATTTGTTCCGCTATTAAGATATAT 360
OY 361 CTGGAATATGTGTCATCCCTCAGACCTTGGGAAAGCCAGGGGCAATGATTAACATAC 420
DB 361 CTGGAATATGTGTCATCCCTCAGACCTTGGGAAAGCCAGGGGCAATGATTAACATAC 420
OY 421 TGGCTGTGTGTGCAATGACCAATGTATGTGTGTAATTTGAGAGCAAGATGTGC 480
DB 421 TGGCTGTGTGTGCAATGACCAATGTATGTGTGTAATTTGAGAGCAAGATGTGC 480
OY 481 AAGAGGTGAAATATGAGGTGAAGTCTAGCATATTGGCTATGATTGAGAGGAGA 540
DB 481 AAGAGGTGAAATATGAGGTGAAGTCTAGCATATTGGCTATGATTGAGAGGAGA 540
OY 541 AACGACTGGAAGTCAATTTGCAATTAATGTGATGATCCTGATGAGCAATTTATATGA 600
DB 541 AACGACTGGAAGTCAATTTGCAATTAATGTGATGATCCTGATGAGCAATTTATATGA 600
OY 601 TATCATATGATGTCAAAAGCGCTGAACCTGTACTTACAGTACTGTGGACTGTTTNG 660
DB 601 TATCATATGATGTCAAAAGCGCTGAACCTGTACTTACAGTACTGTGGACTGTTTNG 660
OY 661 AAGGTATAGGTTCTGTGATGGAACCAAGAAATAGTGTGCTTAATGCAATTTAA 720
DB 661 AAGGTATAGGTTCTGTGATGGAACCAAGAAATAGTGTGCTTAATGCAATTTAA 720
OY 721 AGATTAAGCACTTGGCATTTATTTAAAGCACTATGACATCTGTAAGCAATTTAGT 780
DB 721 AGATTAAGCACTTGGCATTTATTTAAAGCACTATGACATCTGTAAGCAATTTAGT 780
OY 781 GACTAAGAAAGCAATGGAAGAAATCACTGTCATGATATACACTTGTCTGAGAGCC 840
DB 781 GACTAAGAAAGCAATGGAAGAAATCACTGTCATGATATACACTTGTCTGAGAGCC 840
OY 841 CTTCAAGTGTGATCCGATGCTGCGAGACCAATTTGTGATGCTTTTACCAACCTGTGA 900
DB 841 CTTCAAGTGTGATCCGATGCTGCGAGACCAATTTGTGATGCTTTTACCAACCTGTGA 900
OY 901 ATCTGCTGCACAGTACCAAGACGATGATGATGATGATGATGATGATGATGATG 960
DB 901 ATCTGCTGCACAGTACCAAGACGATGATGATGATGATGATGATGATGATGATG 960
OY 961 AATTTCTCTGGAATCAAGCTGATATCTCATCTGATCTGATCTGATCTGATCTG 1020
DB 961 AATTTCTCTGGAATCAAGCTGATATCTCATCTGATCTGATCTGATCTGATCTG 1020
OY 1021 TAAAGTACTAGCTTTTCAAACTTTAAATTTGTAGAACTCATCTAAGTAAATTC 1080
DB 1021 TAAAGTACTAGCTTTTCAAACTTTAAATTTGTAGAACTCATCTAAGTAAATTC 1080
```

```
OY 1081 TGTGTGACTATATCAATATATACAGATGTATCATCTAATTAACATTTTCAATCTCA 1140
DB 1081 TGTGTGACTATATCAATATATACAGATGTATCATCTAATTAACATTTTCAATCTCA 1140
OY 1141 ACTAAGATTAATTTTACACATGCTTAAATATCAACAGATGTATTTGAGTCACTT 1200
DB 1141 ACTAAGATTAATTTTACACATGCTTAAATATCAACAGATGTATTTGAGTCACTT 1200
OY 1201 GTGAATAGATGTGCAAGGGGAGCAGATATGATGTATGTTACATATGTTAGAAAT 1260
DB 1201 GTGAATAGATGTGCAAGGGGAGCAGATATGATGTATGTTACATATGTTAGAAAT 1260
OY 1261 AAAATTAATTTTGTG 1275
DB 1261 AAAATTAATTTTGTG 1275

RESULT 2
US-09-134-593-2
: Sequence 2, Application US/09134593
: Patent No. 5981232
: GENERAL INFORMATION:
: APPLICANT: Hawkins, Phillip R.
: APPLICANT: Hillman, Jennifer L.
: TITLE OF INVENTION: A NOVEL HUMAN PYROPHOSPHATASE
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: California
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: IBM Compatible
: SOFTWARE: FastSeq Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/134,593
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/741,437
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0148 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 855-0555
: TELEFAX: (415) 845-4166
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1275 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: IMMEDIATE SOURCE:
: LIBRARY:
: CLONE: Consensus
: US-09-134-593-2

Query Match 100.0%; Score 1275; DB 2; Length 1275;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CAAGAGTTTNGGGGCTCTCTTGTGAGTGGGCGCGGCTGTGCTGTG 60
DB 1 CAAGAGTTTNGGGGCTCTCTTGTGAGTGGGCGCGGCTGTGCTGTG 60
OY 61 GCAGGGGGGGGGGAGGAGTCCGGCATTATGAGCGGCTTACAGCAGAGGAGCGCGG 120
```


Db 61 GCACGGCGGGGCGAGAGCTCCGACTATGAGCGGCTTCAGCAGCCAGGCGCGCGC 120
OY 121 GCCCTTCCCTGGAGTACCGAGTCTTCCTCAAAAATGAGAAAGACAAATATATATCC 180
Db 121 GCCCTTCCCTGGAGTACCGAGTCTTCCTCAAAAATGAGAAAGACAAATATATATCC 180
OY 181 ATTTTCATGATATTTCAATTTATGAGATAGAGATGTGTTTCATGAGTAGTGAAGTACC 240
Db 181 ATTTTCATGATATTTCAATTTATGAGATAGAGATGTGTTTCATGAGTAGTGAAGTACC 240
OY 241 ACCTGCTCTATATCAAAAATGAGATGTCTACAAAGAGCCCTTTAAACCTATTTAAACA 300
Db 241 ACCTGCTCTATATCAAAAATGAGATGTCTACAAAGAGCCCTTTAAACCTATTTAAACA 300
OY 301 AGATGTGAAAAAAGAAAAACCTCCCTATGTGCGCAATTTGTTCCTCCGATTAAGATATAT 360
Db 301 AGATGTGAAAAAAGAAAAACCTCCCTATGTGCGCAATTTGTTCCTCCGATTAAGATATAT 360
OY 361 CTGGAATATGAGTCCATCCCTCAGACTGGGAGACCCAGGACCAATGATTAACATATAC 420
Db 361 CTGGAATATGAGTCCATCCCTCAGACTGGGAGACCCAGGACCAATGATTAACATATAC 420
OY 421 TGCGTGTGTGAGCAATGACCAATGATGTGTGAATTTGGAAGCAAGTATGTGC 480
Db 421 TGCGTGTGTGAGCAATGACCAATGATGTGTGAATTTGGAAGCAAGTATGTGC 480
OY 481 AAGAGTGAATTAATTTGGCGGGAAGTTCTAGCATATTTGGCTATGATTTGAGAGGGA 540
Db 481 AAGAGTGAATTAATTTGGCGGGAAGTTCTAGCATATTTGGCTATGATTTGAGAGGGA 540
OY 541 AACCGACTGGAAGTCAATTTGCCATTTAATGTGATGATCTGATCAGCCCAATTTAATGA 600
Db 541 AACCGACTGGAAGTCAATTTGCCATTTAATGTGATGATCTGATCAGCCCAATTTAATGA 600
OY 601 TATCAATGATGTCAAAAGGCTGAAACCTGGCTACTTAGAACCTACTGTGAGCTGTTAG 660
Db 601 TATCAATGATGTCAAAAGGCTGAAACCTGGCTACTTAGAACCTACTGTGAGCTGTTAG 660
OY 661 AAGGTATAGGTTCCTGATGGAAGAAACCAAGAAATGAGTTTCGTTAATGAGAAATTTAA 720
Db 661 AAGGTATAGGTTCCTGATGGAAGAAACCAAGAAATGAGTTTCGTTAATGAGAAATTTAA 720
OY 721 AGATTAAGACTTTGCCATTTGATTTAATAAGCACTCATGACCATTTGGAAGCATTTAGT 780
Db 721 AGATTAAGACTTTGCCATTTGATTTAATAAGCACTCATGACCATTTGGAAGCATTTAGT 780
OY 781 GACTTAAGAAAAAGAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 840
Db 781 GACTTAAGAAAAAGAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 840
OY 841 CTTCAAGTGTGATCTGATGCTGCGCAGAGCCATTTGATGATGATGATGATGATGATGATG 900
Db 841 CTTCAAGTGTGATCTGATGCTGCGCAGAGCCATTTGATGATGATGATGATGATGATGATG 900
OY 901 ATCTGCTGACAGTATACCAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATG 960
Db 901 ATCTGCTGACAGTATACCAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATG 960
OY 961 AGATTTCTCTGGAATACAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1020
Db 961 AGATTTCTCTGGAATACAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1020
OY 1021 TAAAGTAGTGTGCTTTCAAGCTTTAATTTGTAAGAACTCATGATTAAGTAAGTAATTC 1080
Db 1021 TAAAGTAGTGTGCTTTCAAGCTTTAATTTGTAAGAACTCATGATTAAGTAAGTAATTC 1080
OY 1081 TGCTGTGATCAATCAATATATCTCAGAAATGTTATCCATCTAAGCAATTTTTCATATCTCA 1140
Db 1081 TGCTGTGATCAATCAATATATCTCAGAAATGTTATCCATCTAAGCAATTTTTCATATCTCA 1140
OY 1141 ACTAAGATTAATTTAGACATGCTTAAATATCAAAAGCAGTTGCTATTTGGAAGTACATT 1200
Db 1141 ACTAAGATTAATTTAGACATGCTTAAATATCAAAAGCAGTTGCTATTTGGAAGTACATT 1200

Db 1141 ACTAAGATTAATTTAGACATGCTTAAATATCAAAAGCAGTTGCTATTTGGAAGTACATT 1200
OY 1201 GTGAATAGATGTGCAAGGAGACATATTTGATGATGATGATGATGATGATGATGATGATG 1260
Db 1201 GTGAATAGATGTGCAAGGAGACATATTTGATGATGATGATGATGATGATGATGATGATG 1260
OY 1261 AAAATTTATTTGCTG 1275
Db 1261 AAAATTTATTTGCTG 1275
RESULT 3
US-08-609-443B-14/C
Sequence 14, Application US/08609443B
Patent No. 5840693
GENERAL INFORMATION:
APPLICANT: ERIKSSON, Ulf
APPLICANT: OLOFSSON, Birgitta
APPLICANT: ALITALO, Kari
APPLICANT: PAISOLA, Kari
TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W., Suite 700
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/609,443B
FILING DATE: 01-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/397,651
FILING DATE: 01-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/469,427
FILING DATE: 06-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/569,063
FILING DATE: 06-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: EVANS, Joseph D
REGISTRATION NUMBER: 26,269
REFERENCE/DOCKET NUMBER: 1064/41979C4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 624 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
TISSUE TYPE: human
US-08-609-443B-14
Query Match 1.4%, Score 18; DB 2; Length 624;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
60 GGCAGCGGCGGCGGCGAG 77
|||||

Db 585 GGCAGCGCGCGCGCAGG 568

RESULT 4

US-08-569-063C-14/c
Sequence 14, Application US/08569063C
Patent No. 5928939

GENERAL INFORMATION:

APPLICANT: ERIKSSON, Ulf
APPLICANT: OLOFSSON, Birgitta
APPLICANT: ALITALO, Kari

APPLICANT: PAJUSOLA, Katri

TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND
TITLE OF INVENTION: DNA CODING THEREFOR

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Evenson, McKeown, Edwards & Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W., Suite 700

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/569,063C

FILING DATE: 06-DEC-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/469,427

FILING DATE: 06-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/397,651

FILING DATE: 01-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: EVANS, Joseph D

REGISTRATION NUMBER: 26,269

REFERENCE/DOCKET NUMBER: 1064/41979CP3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 628-8800

TELEFAX: (202) 628-8844

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 624 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

ORIGINAL SOURCE:

TISSUE TYPE: human

US-08-569-063C-14

Query Match

Best Local Similarity 100.0%; Score 18; DB 2; Length 624;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 585 GGCAGCGCGCGCGCAGG 568

RESULT 5

5459251-1/c

Patent No. 5459251

APPLICANT: Tsujimoto, Yoshida;Croce, Carlo A.

TITLE OF INVENTION: DNA MOLECULES HAVING HUMAN BCL-2 GENE

SEQUENCES

NUMBER OF SEQUENCES: 4

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/228,704

FILING DATE: 18-APR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 994,941

FILING DATE: 23-DEC-1992

APPLICATION NUMBER: 663,010

FILING DATE: 19-MAR-1991

APPLICATION NUMBER: 883,687

FILING DATE: 09-JUL-1986

SEQ ID NO: 1

LENGTH: 4825

Query Match

Best Local Similarity 100.0%; Score 18; DB 6; Length 4825;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 694 TGGCAGCGCGCGCGCAGG 677

Query Match

Best Local Similarity 100.0%; Score 18; DB 2; Length 5086;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 694 TGGCAGCGCGCGCGCAGG 677

RESULT 6

US-08-465-485A-19/c

Sequence 19, Application US/08465485A

Patent No. 5831066

GENERAL INFORMATION:

APPLICANT: Reed, John

TITLE OF INVENTION: Regulation of bcl-2 Gene Expression

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESSES:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, WALTER & NEUSTADT,

ADDRESS: P.C.

STREET: 1755 S. Jefferson Davis Hwy., Suite 400

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/465,485A

FILING DATE: 05-JUN-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/124,256

FILING DATE: 20-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/840,716

FILING DATE: 21-FEB-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/288,692

FILING DATE: 22-DEC-1988

ATTORNEY/AGENT INFORMATION:

NAME: Fortney, Andrew D.

REGISTRATION NUMBER: 34,600

REFERENCE/DOCKET NUMBER: 3335-070-55 CONT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (408) 436-2070

TELEFAX: (408) 436-2075

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 5086 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 TGGCAGCGCGCGCGCAG 76
|||||
Db 694 TGGCAGCGCGCGCGCAG 677

RESULT 7

US-08-365-486A-14/c
Sequence 14, Application US/08365486A
Patent No. 5834306
GENERAL INFORMATION:
APPLICANT: Webster, Keith A.
APPLICANT: Bishopric, Nanette H.
TITLE OF INVENTION: Tissue Specific Hypoxia Regulated
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/365,486A
FILING DATE: 23-DEC-1994
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 8255-0018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 5086 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: human bcl-2 cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1459..2178
US-08-365-486A-14

Query Match 1.4%; Score 18; DB 2; Length 5086;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 TGGCAGCGCGCGCGCAG 76
|||||
Db 694 TGGCAGCGCGCGCGCAG 677

RESULT 8

US-09-080-285-19/c
Sequence 19, Application US/09080285
Patent No. 6040181
GENERAL INFORMATION:
APPLICANT: Reed, John

TITLE OF INVENTION: Regulation of bcl-2 Gene Expression
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: P.C.
STREET: 1755 S. Jefferson Davis Hwy., Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/080,285
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/465,485
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/124,256
FILING DATE: 20-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/840,716
FILING DATE: 21-FEB-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/288,692
FILING DATE: 22-DEC-1988

ATTORNEY/AGENT INFORMATION:
NAME: Fortney, Andrew D.
REGISTRATION NUMBER: 34,600
REFERENCE/DOCKET NUMBER: 3335-070-55 CONT.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (408) 436-2070
TELEFAX: (408) 436-2075
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 5086 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-09-080-285-19

Query Match 1.4%; Score 18; DB 3; Length 5086;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 TGGCAGCGCGCGCGCAG 76
|||||
Db 694 TGGCAGCGCGCGCGCAG 677

RESULT 9

US-08-880-342-14/c
Sequence 14, Application US/08880342
Patent No. 6218179
GENERAL INFORMATION:
APPLICANT: Webster, Keith A.
APPLICANT: Bishopric, Nanette H.
APPLICANT: Murphy, Brian
APPLICANT: Laderoute, Keith R.
APPLICANT: Green, Christopher J.
TITLE OF INVENTION: Tissue Specific Hypoxia Regulated
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto

STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/880,342
FILING DATE: 23-JUN-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/IB95/00996
FILING DATE: 13-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/365,486
FILING DATE: 23-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 8255-0018.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 5086 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: human bcl-2 cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1459..2178
US-08-880-342-14

Query Match 1.4%; Score 18; DB 4; Length 5086;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 TGGCAGCGCGCGCGCAG 76
|||||
DB 694 TGGCAGCGCGCGCGCAG 677

RESULT 10
PCT-US93-05651-4/c
Sequence 4, Application PC/TUS9305651
GENERAL INFORMATION:
TITLE OF INVENTION: A Gene which Prevents Programmed Cell Death
NUMBER OF SEQUENCES: 5
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/05651
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 5086 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1459..2178
OTHER INFORMATION: /product="bcl-2"
PCT-US93-05651-4

Query Match 1.4%; Score 18; DB 5; Length 5086;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 TGGCAGCGCGCGCGCAG 76
|||||
DB 694 TGGCAGCGCGCGCGCAG 677

RESULT 11
PCT-US93-06251-2/c
Sequence 2, Application PC/TUS9306251
GENERAL INFORMATION:
APPLICANT: Wickstrom, Eric and Rife, Jason P.
TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/06251
FILING DATE: 19930630
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8586
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5086 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-2

Query Match 1.4%; Score 18; DB 5; Length 5086;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 TGGCAGCGCGCGCGCAG 76
|||||
DB 694 TGGCAGCGCGCGCGCAG 677

RESULT 12
5506344-1/c
Patent No. 5506344
APPLICANT: TSUJIMOTO, YOSHIHIDE;CROCE, CARLO A.
TITLE OF INVENTION: ANTIBODIES SPECIFIC FOR BCL-2GENE PRODUCT
NUMBER OF SEQUENCES: 5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435,193
FILING DATE: 05-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 228,704

FILING DATE: 18-APR-1994
APPLICATION NUMBER: 994,941
FILING DATE: 23-DEC-1992
APPLICATION NUMBER: 663,010
FILING DATE: 19-MAR-1991
APPLICATION NUMBER: 883,687
FILING DATE: 09-JUL-1986
SEQ ID NO: 1
LENGTH: 5104
5506344-1

Query Match 1.4%; Score 18; DB 6; Length 5104;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 TGGCAGCGCGCGCGCAG 76
Db 694 TGGCAGCGCGCGCGCAG 677

RESULT 13
US-08-180-524-5/c
Sequence 5, Application US/08180524
Patent No. 5849537

GENERAL INFORMATION:
APPLICANT: Tripp, Matthew
APPLICANT: Lusk, Lance
APPLICANT: Rhodes, Thomas
APPLICANT: Huige, Nick
APPLICANT: Kot, Edward
APPLICANT: Chicoye, Etzer
APPLICANT: Barney, Michael C.
APPLICANT: Cronan, Charles L.
TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS
TITLE OF INVENTION: IN YEAST
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Thad Kryshak, Quarles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MSWORD Version 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/180,524
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/917,216
FILING DATE:
APPLICATION NUMBER: US 07/486,333
FILING DATE: 28-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/409,217
FILING DATE: 19-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kryshak, Thad
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 66-005-9234-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5707
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 77 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Oligonucleotide
HYPOTHETICAL: YES
US-08-180-524-5

Query Match 1.3%; Score 17; DB 2; Length 77;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 GTGGCAGCGCGCGCGC 74
Db 54 GTGGCAGCGCGCGCGC 38

RESULT 14
US-08-975-166-5/c
Sequence 5, Application US/08975166
Patent No. 5928877

GENERAL INFORMATION:
APPLICANT: Tripp, Matthew
APPLICANT: Lusk, Lance
APPLICANT: Rhodes, Thomas
APPLICANT: Huige, Nick
APPLICANT: Kot, Edward
APPLICANT: Chicoye, Etzer
APPLICANT: Barney, Michael C.
APPLICANT: Cronan, Charles L.
TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS
TITLE OF INVENTION: IN YEAST
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Thad Kryshak, Quarles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MSWORD Version 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,166
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/180,524
FILING DATE:
APPLICATION NUMBER: US/07/917,216
FILING DATE:
APPLICATION NUMBER: US 07/486,333
FILING DATE: 28-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/409,217
FILING DATE: 19-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kryshak, Thad
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 66-005-9234-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5707
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 77 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Oligonucleotide

HYPOTHETICAL: YES
US-08-975-166-5

Query Match 1.3%; Score 17; DB 2; Length 77;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 GTGGCAGCGCGCGCGC 74
|||||
DB 54 GTGGCAGCGCGCGCGC 38

RESULT 15

US-08-180-524-3/C
Sequence 3, Application US/08180524
Patent No. 5849537

GENERAL INFORMATION:

APPLICANT: Tripp, Matthew

APPLICANT: Lusk, Lance

APPLICANT: Rhodes, Thomas

APPLICANT: Huige, Nick

APPLICANT: Kot, Edward

APPLICANT: Chicoye, Elzer

APPLICANT: Barney, Michael C.

APPLICANT: Bower, Patricia A.

APPLICANT: Cronan, Charles L.

TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS

TITLE OF INVENTION: IN YEAST

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Thad Kryshak, Quarles & Brady

STREET: 411 East Wisconsin Avenue

CITY: Milwaukee

STATE: Wisconsin

COUNTRY: USA

ZIP: 53202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MSWORD Version 5.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/180.524

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/917.216

FILING DATE:

APPLICATION NUMBER: US 07/486.333

FILING DATE: 28-FEB-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/409.217

FILING DATE: 19-SEP-1989

ATTORNEY/AGENT INFORMATION:

NAME: Kryshak, Thad

REGISTRATION NUMBER: 35,433

REFERENCE/DOCKET NUMBER: 66-005-9234-1

TELEPHONE: (414) 277-5707

TELEFAX: (414) 277-5591

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 146 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: Other Nucleic Acid

HYPOTHETICAL: YES

ANTI-SENSE: NO

US-08-180-524-3

Query Match 1.3%; Score 17; DB 2; Length 146;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 GTGGCAGCGCGCGCGC 74
|||||
DB 123 GTGGCAGCGCGCGCGC 107

RESULT 16

US-08-975-166-3/C
Sequence 3, Application US/08975166
Patent No. 5928877

GENERAL INFORMATION:

APPLICANT: Tripp, Matthew

APPLICANT: Lusk, Lance

APPLICANT: Rhodes, Thomas

APPLICANT: Huige, Nick

APPLICANT: Kot, Edward

APPLICANT: Chicoye, Elzer

APPLICANT: Barney, Michael C.

APPLICANT: Bower, Patricia A.

APPLICANT: Cronan, Charles L.

TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS

TITLE OF INVENTION: IN YEAST

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Thad Kryshak, Quarles & Brady

STREET: 411 East Wisconsin Avenue

CITY: Milwaukee

STATE: Wisconsin

COUNTRY: USA

ZIP: 53202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MSWORD Version 5.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/975.166

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/180.524

FILING DATE:

APPLICATION NUMBER: US/07/917.216

FILING DATE:

APPLICATION NUMBER: US 07/486.333

FILING DATE: 28-FEB-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/409.217

FILING DATE: 19-SEP-1989

ATTORNEY/AGENT INFORMATION:

NAME: Kryshak, Thad

REGISTRATION NUMBER: 35,433

REFERENCE/DOCKET NUMBER: 66-005-9234-1

TELEPHONE: (414) 277-5707

TELEFAX: (414) 277-5591

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 146 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: Other Nucleic Acid

HYPOTHETICAL: YES

ANTI-SENSE: NO

US-08-975-166-3

Query Match 1.3%; Score 17; DB 2; Length 146;
Best Local Similarity 100.0%; Pred. No. 35;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 58 GTGGACGGCGCGGC 74
|||||
Db 123 GTGGACGGCGCGGC 107

RESULT 17

US-08-454-720A-37/C
Sequence 37, Application US/08454720A
Patent No. 576688

GENERAL INFORMATION:

APPLICANT: Sobol, Robert E.
APPLICANT: Green, Mark R.
APPLICANT: Kawasaki, Ernest S.
TITLE OF INVENTION: Detection of Carcinoma Metastases by
TITLE OF INVENTION: Nucleic Acid Amplification
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffman-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.01, version
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/454,720A
FILING DATE: May 31, 1995
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/357,565
FILING DATE: December 16, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Petry, Douglas
REGISTRATION NUMBER: 35321
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2974
TELEFAX: (510) 814-2977
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 461 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-454-720A-37

Query Match 1.3%; Score 17; DB 1; Length 461;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 58 GTGGACGGCGCGGC 74
|||||
Db 36 GTGGACGGCGCGGC 20

RESULT 18

US-09-328-111-577
Sequence 577, Application US/09328111
Patent No. 6262333

GENERAL INFORMATION:

APPLICANT: Endege, Wilson O.
APPLICANT: Steinmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.

APPLICANT: Carroll III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Dertl, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
TITLE OF INVENTION: PRODUCTS
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328,111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 577
LENGTH: 617
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(617)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-577

Query Match 1.3%; Score 17; DB 4; Length 617;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 511 AGGCATATGGCTATGA 527
|||||
Db 470 aggcatttgctatga 486

RESULT 19

US-09-154-083-31
Sequence 31, Application US/09154083
Patent No. 6150513

GENERAL INFORMATION:

APPLICANT: Wu, Kai
TITLE OF INVENTION: Polypeptide Synthase Enzymes and Recombinant DNA
TITLE OF INVENTION: Constructs therefor
FILE REFERENCE: 30062-20022.00 Polypeptide Synthase Enz
CURRENT APPLICATION NUMBER: US/09/154,083
CURRENT FILING DATE: 1998-09-16
NUMBER OF SEQ ID NOS: 31
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO: 31
LENGTH: 718
TYPE: DNA
ORGANISM: Streptomyces hygroscopicus
US-09-154-083-31

Query Match 1.3%; Score 17; DB 3; Length 718;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 104 ACCGAGAGCGCGCCGC 120
|||||
Db 204 accgagagcgccgcgc 220

RESULT 20

US-09-097-072-1/C
Sequence 1, Application US/09097072
Patent No. 6140079

GENERAL INFORMATION:

APPLICANT: Kallender, Howard
APPLICANT: Lennox, Anna L.
APPLICANT: Burnham, Martin K. R.

APPLICANT: Ward, Judith
TITLE OF INVENTION: 91dB
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert, Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2793
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/097,072
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/886,638
FILING DATE: 01-JUL-1997
ATTORNEY/AGENT INFORMATION:
NAME: Falk, Stephen T
REGISTRATION NUMBER: 36,795
REFERENCE/DOCKET NUMBER: GM10028-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 810 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-097-072-1

Query Match 1.3%; Score 17; DB 3; Length 810;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 991 TACATCGTTCATCTG 1007
|||||
DB 203 TACATCGTTCATCTG 187

RESULT 21
US-08-886-638-1/c
Sequence 1, Application US/08886638
Patent No. 5866367
GENERAL INFORMATION:
APPLICANT: Kallender, Howard
TITLE OF INVENTION: NOVEL COMPOUNDS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert Price & Rhoads
STREET: 997 Lenox Drive, Building 3, Suite 210
CITY: Lawrenceville
STATE: NJ
COUNTRY: USA
ZIP: 08543
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/886,638
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bloom, Allen
REGISTRATION NUMBER: 29,135
REFERENCE/DOCKET NUMBER: GM10028
TELECOMMUNICATION INFORMATION:
TELEPHONE: 609-520-3214
TELEFAX: 609-520-3259
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 900 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-886-638-1

Query Match 1.3%; Score 17; DB 2; Length 900;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 991 TACATCGTTCATCTG 1007
|||||
DB 221 TACATCGTTCATCTG 205

RESULT 22
US-09-097-072-3/c
Sequence 3, Application US/09097072
Patent No. 6140079
GENERAL INFORMATION:
APPLICANT: Kallender, Howard
APPLICANT: Lennox, Anna L.
APPLICANT: Burdham, Martin K. R.
TITLE OF INVENTION: 91dB
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert, Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2793
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/097,072
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/886,638
FILING DATE: 01-JUL-1997
ATTORNEY/AGENT INFORMATION:
NAME: Falk, Stephen T
REGISTRATION NUMBER: 36,795
REFERENCE/DOCKET NUMBER: GM10028-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 900 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-097-072-3

Query Match 1.3%: Score 17; DB 3; Length 900;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 991 TACATCGTGTTCATCTG 1007
|||||
DB 221 TACATCGTGTTCATCTG 205

RESULT 23

US-09-212-988-1/c
Sequence 1, Application US/09212988
Patent No. 6210673
GENERAL INFORMATION:
APPLICANT: Kallender, Howard
TITLE OF INVENTION: NOVEL COMPOUNDS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert Price & Rhoads
STREET: 997 Lenox Drive, Building 3, Suite 210
CITY: Lawrenceville
STATE: NJ
COUNTRY: USA
ZIP: 08543
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/212.988
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/886,638
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bloom, Allen
REGISTRATION NUMBER: 29,135
REFERENCE/DOCKET NUMBER: GM10028
TELECOMMUNICATION INFORMATION:
TELEPHONE: 609-520-3214
TELEFAX: 609-520-3259
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 900 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-212-988-1

Query Match 1.3%: Score 17; DB 4; Length 900;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 991 TACATCGTGTTCATCTG 1007
|||||
DB 221 TACATCGTGTTCATCTG 205

RESULT 24
PCT-US91-00899-13
Sequence 13, Application PC/TUS9100899
GENERAL INFORMATION:
APPLICANT: Lowe, John B.
TITLE OF INVENTION: Method and Products for the Synthesis of
Oligosaccharide Structures on Glycoproteins, Glycolipids,
or as Free Molecules, and for the Isolation of Cloned
Genetic Sequences That Determine These Structures

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

ADDRESSEE: P.C. Jefferson Davis Highway, Suite 400

STREET: Arlington

CITY: Arlington

STATE: Virginia

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US91/00899

FILING DATE: 19910214

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye Ph.D., Jean-Paul

REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 2363-021-55 PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)521-5940

TELEFAX: (703)486-2347

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 1068 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

PCT-US91-00899-13

Query Match 1.3%: Score 17; DB 5; Length 1068;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 GGCACGGCGGCGGCAG 76
|||||
DB 294 GGCACGGCGGCGGCAG 310

RESULT 25

PCT-US91-00899-5
Sequence 5, Application PC/TUS9100899
GENERAL INFORMATION:
APPLICANT: Lowe, John B.
TITLE OF INVENTION: Method and Products for the Synthesis of
Oligosaccharide Structures on Glycoproteins, Glycolipids,
or as Free Molecules, and for the Isolation of Cloned
Genetic Sequences That Determine These Structures
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/00899
FILING DATE: 19910214
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye Ph.D., Jean-Paul

REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-021-55 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-5940
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1215 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
PCT-US91-00899-5

Query Match 1.3%; Score 17; DB 5; Length 1215;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 GGCAGCGCGCGCGCAG 76
|||||
DB 441 GGCAGCGCGCGCGCAG 457

RESULT 26
US-07-914-281-9
Sequence 9, Application US/07914281
Patent No. 5324663

GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.,
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/914,281
FILING DATE: 19920720
CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M. P.

REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1488 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)
US-07-914-281-9

Query Match 1.3%; Score 17; DB 1; Length 1488;
Best Local Similarity 100.0%; Pred. No. 39;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 60 GGCAGCGCGCGCGCAG 76
|||||
DB 714 GGCAGCGCGCGCGCAG 730

RESULT 27
US-08-393-246-9
Sequence 9, Application US/08393246
Patent No. 5595900

GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.,
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/393,246
FILING DATE:
CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/220,433
FILING DATE: 30-MAR-1994
APPLICATION NUMBER: US 07/914,281
FILING DATE: 20-JUL-1992

ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1488 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)
US-08-393-246-9

Query Match 1.3%; Score 17; DB 1; Length 1488;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 GGCAGCGCGCGCGCAG 76
|||||
DB 714 GGCAGCGCGCGCGCAG 730

RESULT 28
US-08-525-058A-9

Sequence 9, Application US/08525058A
Patent No. 5770420
GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.

TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P. C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/525,058A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lavallee, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1488 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-525-058A-9

Query Match 1.38; Score 17; DB 1; Length 1488;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 60 GGCAGCGGGCGGCAG 76
DB 714 GGCAGCGGGCGGCAG 730

RESULT 29
US-08-696-731-9
Sequence 9, Application US/08696731
Patent No. 595347
GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P. C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/696,731
FILING DATE: 14-AUG-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/393,246
FILING DATE:
APPLICATION NUMBER: US 08/220,433
FILING DATE: 30-MAR-1994
APPLICATION NUMBER: US 07/914,281
FILING DATE: 20-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lavallee, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1488 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-696-731-9

Query Match 1.38; Score 17; DB 2; Length 1488;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 60 GGCAGCGGGCGGCAG 76
DB 714 GGCAGCGGGCGGCAG 730

RESULT 30
US-09-042-531-9
Sequence 9, Application US/09042531
Patent No. 6268193
GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P. C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,531
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/393,246
FILING DATE:
APPLICATION NUMBER: US 08/220,433
FILING DATE: 30-MAR-1994
APPLICATION NUMBER: US 07/914,281
FILING DATE: 20-JUL-1992
ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAR UR
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1488 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-09-042-531-9

Query Match 1.3%; Score 17; DB 4; Length 1488;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 GGCAGCGCGCGCGCGAG 76
|||||
Db 714 GGCAGCGCGCGCGCGAG 730

RESULT 31
US-08-365-486A-16/c
Sequence 16, Application US/08365486A
Patent No. 5834306
GENERAL INFORMATION:
APPLICANT: Webster, Keith A.
APPLICANT: Bishopric, Nanette H.
TITLE OF INVENTION: Tissue Specific Hypoxia Regulated
TITLE OF INVENTION: Therapeutic Constructs
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESS: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/365,486A
FILING DATE: 23-DEC-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 8255-0018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1846 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: bcl-2 fusion gene; Seto, et al.,
INDIVIDUAL ISOLATE: EMBO J 7:123 (1988)
FEATURE:
NAME/KEY: CDS

LOCATION: 887..1606
US-08-365-486A-16

Query Match 1.3%; Score 17; DB 2; Length 1846;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 GGCAGCGCGCGCGCGAG 76
|||||
Db 339 GGCAGCGCGCGCGCGAG 323

RESULT 32
US-08-880-342-16/c
Sequence 16, Application US/08880342
Patent No. 6218179
GENERAL INFORMATION:
APPLICANT: Webster, Keith A.
APPLICANT: Bishopric, Nanette H.
APPLICANT: Murphy, Brian
APPLICANT: Laderoute, Keith R.
APPLICANT: Green, Christopher J.
TITLE OF INVENTION: Tissue Specific Hypoxia Regulated
TITLE OF INVENTION: Therapeutic Constructs
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESS: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/880,342
FILING DATE: 23-JUN-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/IB95/00996
FILING DATE: 13-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/365,486
FILING DATE: 23-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 8255-0018.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1846 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: bcl-2 fusion gene; Seto, et al.,
INDIVIDUAL ISOLATE: EMBO J 7:123 (1988)
FEATURE:
NAME/KEY: CDS
LOCATION: 887..1606
US-08-880-342-16

Query Match 1.38; Score 17; DB 4; Length 1846;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 GGCAGCGCGCGCGCAG 76
Db 339 GGCAGCGCGCGCGCAG 323

RESULT 33

US-08-483-151-3
Sequence 3, Application US/08483151
Patent No. 5858752
GENERAL INFORMATION:
APPLICANT: Seed, Brian
ATTORNEY/AGENT INFORMATION:
TITLE OF INVENTION: FUCOSYLTRANSFERASE GENES AND USES THEREOF
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,151
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Lech, Karen F.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/278001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2134 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-483-151-3

Query Match 1.38; Score 17; DB 2; Length 2134;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 GGCAGCGCGCGCGCAG 76
Db 462 GGCAGCGCGCGCGCAG 478

RESULT 34

PCT-US96-06427-3
Sequence 3, Application PC/TUS9606427
GENERAL INFORMATION:
APPLICANT: The General Hospital Corporation
TITLE OF INVENTION: FUCOSYLTRANSFERASE GENES AND USES THEREOF
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA

COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/06427
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/483,151
FILING DATE: 07-JUN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Lech, Karen F.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/278001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2134 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
PCT-US96-06427-3

Query Match 1.38; Score 17; DB 5; Length 2134;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 GGCAGCGCGCGCGCAG 76
Db 462 GGCAGCGCGCGCGCAG 478

RESULT 35

US-08-482-073-9
Sequence 9, Application US/08482073
Patent No. 6307025
GENERAL INFORMATION:
APPLICANT: Hession, Catherine A.
APPLICANT: Lobd, Roy R.
APPLICANT: Goetz, Susan E.
APPLICANT: Osborn, Laurelee
APPLICANT: Benjamin, Christopher D.
TITLE OF INVENTION: ENDOTHELIAL CELL-LEUKOCYTE ADHESION
TITLE OF INVENTION: MOLECULES (ELAMS) AND MOLECULES INVOLVED IN LEUKOCYTE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,073
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/486,336
FILING DATE: 31-OCT-1990
APPLICATION NUMBER: US 07/608298
FILING DATE: 31-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US 90/02357
FILING DATE: 27-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/452675
FILING DATE: 18-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/359516
FILING DATE: 01-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/354151
FILING DATE: 28-APR-1989
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: B124CIP4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
TELEX: 14-8367
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2175 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-482-073-9

Query Match 1.3%; Score 17; DB 4; Length 2175;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 60 GGCAGCGCGCGCGCGCAG 76
|||||
DB 506 GGCAGCGCGCGCGCGCAG 522

RESULT 36
US-08-482-073-10
Sequence 10, Application US/08482073
Patent No. 6307025
GENERAL INFORMATION:
APPLICANT: Hession, Catherine A.
APPLICANT: Lobb, Roy R.
APPLICANT: Goelz, Susan E.
APPLICANT: Osborn, Laurelee
APPLICANT: Benjamin, Christopher D.
APPLICANT: Rosa, Margaret D.
TITLE OF INVENTION: ENDOTHELIAL CELL-LEUKOCYTE ADHESION
TITLE OF INVENTION: MOLECULES (ELAMS) AND MOLECULES INVOLVED IN LEUKOCYTE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,073
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/486,336
FILING DATE: 31-OCT-1990
APPLICATION NUMBER: US 07/608298
FILING DATE: 31-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US 90/02357
FILING DATE: 27-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/452675
FILING DATE: 18-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/359516
FILING DATE: 01-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/354151
FILING DATE: 28-APR-1989
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: B124CIP4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
TELEX: 14-8367
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2861 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-482-073-10

Query Match 1.3%; Score 17; DB 4; Length 2861;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 60 GGCAGCGCGCGCGCGCAG 76
|||||
DB 989 GGCAGCGCGCGCGCGCAG 1005

RESULT 37
US-08-006-676B-2
Sequence 2, Application US/08006676B
Patent No. 5411865
GENERAL INFORMATION:
APPLICANT: Reed, Steven
TITLE OF INVENTION: Diagnosis of Leishmaniasis
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jeffrey B. Oster
STREET: 8339 SE 57th Street
CITY: Mercer Island
STATE: Washington
COUNTRY: USA
ZIP: 98040-4906
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORD for Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/006,676B
FILING DATE: 15-JAN-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Oster, Jeffrey B.
REGISTRATION NUMBER: 32,585
REFERENCE/DOCKET NUMBER: REED-4
TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 232 7845
TELEFAX: (206) 236 0205
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3319 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE: Leishmania chagasi
US-08-006-676B-2

Query Match 1.3%; Score 17; DB 1; Length 3319;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 CCGAGAGCGCGCGCG 121
|||||
Db 3288 CCGAGAGCGCGCGCG 3304

RESULT 38
US-08-282-845-1
Sequence 1, Application US/08282845
Patent No. 5719263
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: A 230kd Antigen Present in Leishmania
TITLE OF INVENTION: Species
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple Macintosh Operating System 7.1
SOFTWARE: Microsoft Word for Macintosh 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/282,845
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/006,676
FILING DATE: JANUARY 15, 1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: 5004-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)233-0644
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3319 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: rK39
FEATURE:
NAME/KEY: CDS
LOCATION: 455..3319

US-08-282-845-1

Query Match 1.3%; Score 17; DB 1; Length 3319;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 CCGAGAGCGCGCGCG 121
|||||
Db 3288 CCGAGAGCGCGCGCG 3304

RESULT 39
US-08-428-414A-4
Sequence 4, Application US/08428414A
Patent No. 5912166
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TITLE OF INVENTION: LEISHMANIASIS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,414A
FILING DATE: 21-APR-1995
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Kadlec, Ann T.
REGISTRATION NUMBER: 39,244
REFERENCE/DOCKET NUMBER: 210121.407
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANDBERRY
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3319 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
US-08-428-414A-4

Query Match 1.3%; Score 17; DB 2; Length 3319;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 CCGAGAGCGCGCGCG 121
|||||
Db 3288 CCGAGAGCGCGCGCG 3304

RESULT 40
PCT-US94-00324-2
Sequence 2, Application PC/TUS9400324
GENERAL INFORMATION:
APPLICANT: Reed, Steven
TITLE OF INVENTION: Diagnosis of Leishmaniasis
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation

STREET: 51 University Street
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple System 7.1
SOFTWARE: Microsoft Word, version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/00324
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/006,676
FILING DATE: 15-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: 5004-WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3319 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Leishmania chagasi
PCT-US94-00324-2

Query Match 1.3%; Score 17; DB 5; Length 3319;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 CCGAGGAGCGCGCGCG 121
DB 3288 CCGAGGAGCGCGCGCG 3304

RESULT 41
US-08-972-719-1/c
Sequence 1, Application US/08972719
Patent No. 6060588
GENERAL INFORMATION:
APPLICANT: LI, SHENG FENG
APPLICANT: PHILLIPS, DAVID
TITLE OF INVENTION: IDENTIFICATION OF Bap-1, A PROTEIN THAT
TITLE OF INVENTION: BINDS TO INTEGRIN AND IS INVOLVED IN INTEGRIN-MEDIATED
TITLE OF INVENTION: SIGNAL TRANSDUCTION
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 PENNSYLVANIA AVENUE, NW
CITY: WASHINGTON
STATE: CA
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/972,719
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/753,038
FILING DATE: 19-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: MILLMAN, ROBERT A.
REGISTRATION NUMBER: 36,217
REFERENCE/DOCKET NUMBER: 22803-20011.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 822-0168
TELEX: 90-4030 MRSNFOERSMWH
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3467 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 226..1233
US-08-972-719-1

Query Match 1.3%; Score 17; DB 3; Length 3467;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1233 ATGTATATGTATACATA 1249
DB 3379 ATGTATATGTATACATA 3363

RESULT 42
US-08-753-038-1/c
Sequence 1, Application US/08753038
Patent No. 6194557
GENERAL INFORMATION:
APPLICANT: LI, SHENG FENG
APPLICANT: PHILLIPS, DAVID
TITLE OF INVENTION: IDENTIFICATION OF Bap-1, A PROTEIN THAT
TITLE OF INVENTION: BINDS TO INTEGRIN AND IS INVOLVED IN INTEGRIN-MEDIATED
TITLE OF INVENTION: SIGNAL TRANSDUCTION
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 PENNSYLVANIA AVENUE, NW
CITY: WASHINGTON
STATE: CA
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/753,038
FILING DATE: 19-NOV-1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: MILLMAN, ROBERT A.
REGISTRATION NUMBER: 36,217
REFERENCE/DOCKET NUMBER: 22803-20011.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 822-0168
TELEX: 90-4030 MRSNFOERSMWH
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3467 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 226..1233
US-08-753-038-1

Query Match
Best local Similarity 1.3%; Score 17; DB 4; Length 3647;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1233 ATGTATATGTTACCATA 1249
|||||
DB 3379 ATGTATATGTTACCATA 3363

RESULT 43
US-07-914-281-7

Sequence 7, Application US/07914281
Patent No. 5324663

GENERAL INFORMATION:

APPLICANT: LOWE, JOHN B.

TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS

TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,

TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION

TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,

ADDRESS: P.C.

STREET: 1755 Jefferson Davis Highway, Fourth Floor

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/914,281

FILING DATE: 19920720

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye, Jean-Paul M. P.

REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 2363-060-55

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)521-4500

TELEFAX: (703)486-2347

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 3647 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

ANTI-SENSE: NO

US-07-914-281-7

Query Match

Best local Similarity 1.3%; Score 17; DB 1; Length 3647;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 GGCAGCGCGCGCGCAG 76
|||||

DB 2382 GGCAGCGCGCGCGCAG 2398

RESULT 44

US-08-393-246-7
Sequence 7, Application US/08393246
Patent No. 5595900

GENERAL INFORMATION:

APPLICANT: LOWE, JOHN B.

TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS

TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,

TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION

TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,

ADDRESS: P.C.

STREET: 1755 Jefferson Davis Highway, Fourth Floor

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/393,246

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/220,433

FILING DATE: 30-MAR-1994

APPLICATION NUMBER: US 07/914,281

FILING DATE: 20-JUL-1992

ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye, Jean-Paul M. P.

REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 2363-060-55

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)521-4500

TELEFAX: (703)486-2347

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 3647 base pairs

TYPE: nucleic acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

ANTI-SENSE: NO

US-08-393-246-7

Query Match

Best local Similarity 1.3%; Score 17; DB 1; Length 3647;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 GGCAGCGCGCGCGCAG 76
|||||

DB 2382 GGCAGCGCGCGCGCAG 2398

RESULT 45

US-08-525-058A-7

Sequence 7, Application US/08525058A

Patent No. 5770420

GENERAL INFORMATION:

APPLICANT: LOWE, JOHN B.

TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS

TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,

TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION

TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT, P.C.

STREET: 1755 Jefferson Davis Highway, Fourth Floor
 CITY: Arlington
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/525,058A
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Lavalleye, Jean-Paul M. P.
 REGISTRATION NUMBER: 31,451
 REFERENCE/DOCKET NUMBER: 2363-060-55
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703)521-4500
 TELEFAX: (703)486-2347
 TELEX: 248855 OPAT UR
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3647 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ANTI-SENSE: NO
 US-08-525-058A-7

Query Match 1.38; Score 17; DB 1; Length 3647;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 60 GGCAGCGGGGGCGGCGAG 76
 ||||||||||||||||
 DB 2382 GGCAGCGGGGGCGGCGAG 2398

Search completed: March 29, 2002, 20:53:54
 Job time: 3524 sec